Brain-Wave Measures of Workload in Advanced Cockpits: The Transition of Technology From Laboratory to Cockpit Simulator

Richard L. Horst, David L. Mahaffey,
and Robert C. Munson

NASA

Brain-Wave Measures of Workload in Advanced Cockpits: The Transition of Technology From Laboratory to Cockpit Simulator

Richard L. Horst, David L. Mahaffey,
and Robert C. Munson
Applied Behavioral Research Group
Advanced Resource Development Corporation
Columbia, Maryland

PREFACE

This report, submitted to NASA by ARD Corporation as a deliverable under Contract NAS1-18019, is the final report for the Phase II Small Business Innovative Research (SBIR) Contract entitled "Brain-wave Measures of Workload in Advanced Cockpits." The Contract Officer's Technical Representative was Dr. Alan Pope of the Crew/Vehicle Interface Research Branch, NASA Langley Research Center. Dr. Randall Harris of the same organization also monitored the contract during part of its three-year duration. We greatly appreciate their guidance and support of our efforts.

We also wish to acknowledge the support and assistance of a number of other people on various aspects of the present project:

o   The empirical studies were conducted by ARD personnel using Dr. Daniel S. Ruchkin's laboratory at the University of Maryland School of Medicine. Dr. Ruchkin, acting as a consultant to ARD, offered valuable guidance in data analysis. He also supported the single-trial analysis task, by designing the software that generated a simulated electroencephalogram (EEG) with embedded Event-related Potentials (ERPs), and the data analysis software upgrade, by offering insights into the functioning of the existing code.

o   Mr. Kemper Kibler and Mr. William Kahlbaum of NASA Langley and Mr. Myron Sothcott of Unisys Corporation, an on-site support contractor at NASA Langley, provided us with valuable information about the capabilities of the NASA Advanced Concepts Flight Simulator (ACFS). They assisted us in defining flight scenarios for eventual use in a simulator-based validation of Event-Related Potential (ERP) measures of workload and in designing communication protocols for passing information between the ACFS computer and other laboratory data collection systems.

o   Mr. George Sexton of Lockheed-Georgia Corporation, acting as a consultant under Lockheed's LEND program, played a key role in our analysis of the NASA Advanced Concepts Flight Simulator (ACFS) regarding cockpit events that are likely to elicit ERPs related to workload. Mr. Sexton explained the design of the "baseline" ACFS displays to us and offered guidance in the definition of appropriate flight scenarios for use in validating our

iii

findings.

o Dr. Richard Shannon of ARD worked closely with Mr. Sexton in the analysis of the ACFS and was responsible for defining the flight scenario and likely use of displays which resulted from this phase of the project.

o Dr. Jorge Aunon of Colorado State University generously provided us with a copy of the public domain FORTRAN software that he and Dr. Clare McGillem developed at Purdue University to implement their approach to single-trial ERP analysis. Acting as a consultant, Dr. Aunon also provided useful guidance to us in understanding the approach that underlies this software.

o Mr. Brian Foote of the Cognitive Psychophysiology Laboratory (CPL), University of Illinois. working as a consultant to ARD, developed the LABPAK software routines and associated documentation that were implemented on the MINC computer at NASA Langley in preparation for the simulator-based validation study. These routines were made available to ARD with the permission of Dr. Emanuel Donchin, Director of the CPL. A more extensive version of the LABPAK routines has been implemented on the PEARL system developed by the CPL and is proprietary to that group.

o Mr. Dan T. Smith, a consultant to ARD, contributed substantially to the single-trial analysis study. He was responsible for many of the ideas embodied in the approach we followed in implementing the Woody and Aunon/McGillem techniques. He also coded the routines that generated simulated data, as well as the program that implemented the comparative single-trial analysis approach that we espouse.

o Mr. C. Kenneth Bond of ARD designed and coded the data management "shell" program which expedited the single-trial analyses that we performed. He also contributed to the design and implementation of the graphics features in the upgrade of our data analysis software.

o Ms. Anne Francoeur of ARD supported the collection and analysis of empirical data.

o Mr. Kevin Quinn of ARD contributed to the upgrade of our data analysis software, including the initial implementation of graphics features to support interactive data analysis.

o Mr. J. B. Winter of ARD supported the analysis of ACFS displays and procedures.

Much of the work summarized here has been previously presented in separate

reports submitted to NASA in the course of the present project. In particular, the laboratory-based empirical studies (Chapter 2) have been discussed in the report "Event-related Potential Indicants of Mental Workload, Attention, and Target Recognition when Monitoring a Multi-Element Display"; the task analysis of the ACFS (Chapter 3) has been documented in the report, "Analysis of Advanced Concepts Flight Simulator Displays and Aircrew Tasks to Identify Events Likely to Elicit Brain-Wave Correlates of Mental Workload"; the ANALYZ software (Chapter 4) has been supplied to NASA (with SBIR rights) along with an "ANALYZ User's Manual" and an "ANALYZ Programmer's Reference Manual"; the single-trial analysis software (Chapter 5) has been delivered to NASA (with SBIR rights) along with "Documentation for Event-related Potential Single-trial Analysis Programs" and has been discussed in the report "An Approach to the Comparative Study of Alternative Methods of Event-related Potential Single-trial Analysis"; the Labpack software (Chapter 6) has been delivered to NASA (with SBIR rights) along with "Documentation for LABPACK Subroutines"; and the specifications for a workload assessment system (Chapter 7) have been submitted (with SBIR rights) in the report, "Design Specification for a General-Purpose Mental Workload Assessment System."

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

Advances in technology have allowed man-in-the-loop control systems to become more complex and automated. With these changes have come changes in the nature of man-machine interactions. The tasks required of the human operators of such systems place emphasis on supervisory skills rather than on the visual-motor skills required to manually control the system. The cognitive functions required of the system supervisor entail such processes as the ability to distribute attention among competing inputs, to integrate large amounts of information, to detect trends among interacting variables which indicate system status, and to make sound decisions in the face of uncertainty. Nowhere are these design trends and cognitive challenges more apparent than in recent and near-term aircraft and spacecraft. Aircrews increasingly share tasks with on-board computers and function as monitors of system performance, intervening when automated systems fail or when conditions are encountered which require human judgement and problem-solving.

The human factors problems that are inherent in such advanced control systems span two extremes. Under high mental workload conditions, imposed by on-board emergencies or environmental threats, there can be an overload of information impinging on the crew. Under low workload conditions, when automated systems are functioning properly and conditions are routine, the human operator can lapse into inattentiveness and is ill-prepared to deal with emergencies. For the full range of conditions bounded by these extremes, the effective and timely transmission of information from system to operator is critical. It is important, therefore, in the system design process to take mental workload into account. This orientation requires that there be valid and reliable measures of mental workload available to the system designer. Several recent review articles document the scientific progress that is being made on this front (Moray, 1979; Hart, 1987; Gopher & Donchin, 1987).

1.1 Usefulness of Mental Workload Measures

Operator effectiveness is ultimately defined in terms of behavioral output.

However, there seems to be both diagnostic and prescriptive value in attempting to develop measures of mental workload, rather than focusing just on observable task performance. For example, task performance may deteriorate for a wide variety of reasons. An operator may miss an alarm signal either because he was cognitively overloaded or because he was bored and not sufficiently vigilant. A system designer, or co-pilot, would take different remedial actions, depending on which of these "mental states" led to the degradation in performance. Furthermore, many task environments allow the human operator to function with some spare capacity such that, to some extent, increased task demands can be met with increased effort in order to maintain behavioral output at a relatively constant level. In such situations, mental workload indices may predict susceptibility to an impending deterioration in performance, should task demands increase still further. Finally, when task demands are low, there may be little behavioral output from which one can gauge the status of the operator. A sense of the operator's mental workload in such situations could be used to infer whether or not such lack of responding was appropriate and the extent to which the operator is prepared to respond appropriately should conditions change. Therefore, the diagnostic and, hopefully, prescriptive value of the mental workload construct is somewhat akin to that of clinical syndromes. Analogous to the different treatments which may be prescribed depending on a clinical diagnosis, inferences about the mental workload which underlies an observed performance deficit may suggest alternative design or operational "treatments."

The danger in using mental workload conceptualizations to explain data, of course, lies in our tendency to think that if we can label something, we have understood it. Terms like "boredom" may not imply the same "syndrome" to everyone. Therefore, until we have sufficient data to define what are the distinguishing features and performance-related consequences of "boredom," it is imperative that we continue to operationally define our use of such terms.

1.2   The Value of Physiological Measures of Workload

Regardless of the stock one puts in the explanatory power of mental workload, it follows from the above discussion that it would be unwise to evaluate and predict an operator's ability to perform solely from observing behavior on a

primary task. Performance on secondary tasks can be instructive for measuring the processing capacity entailed by a primary task. However, with this approach it is difficult to ensure that the operator always gives mental priority to the primary task, the results may be of questionable validity if used to generalize to situations in which the primary task is performed alone, and incompatibilities between the behavioral responses required by the two tasks may make it difficult to draw inferences about the demands placed on perceptual or decision-making processes. Moreover, the sort of contrived secondary tasks that have often been used in laboratory studies are clearly not acceptable in operational settings, so secondary task measures must be found among the activities that the operator is doing in the course of normal operations.

Simply asking the operator for subjective ratings of his perceived state is often useful, but is also fraught with difficulties. The operator may not realize that his environmentally-defined workload is high when, in fact, it is. Furthermore, such subjective ratings tend to be unreliable when administered in operational settings while the operator is simultaneously trying to maintain task performance, and the mere act of completing the rating itself, of course, constitutes an additional task burden on the operator.

For these reasons, there is considerable appeal to the prospects of gaining additional information about the functional status of operators from their physiological signs. Much evidence now suggests (see Horst, 1988) that, if interpreted in conjunction with behavioral and subjective measures, physiological indices offer the possibility of objectively inferring, not only the general physical fitness to perform, but also the cognitive status of an operator. Physiological measures can often be used to confirm the conclusions derived from behavioral or subjective measures. There are also instances in the literature of physiological measures providing complementary information regarding cognitive activity to that which is available from behavioral measures.

1.3 ERP Measures of Workload

Transient ERPs are usually extracted from the ongoing EEG by signal averaging

over numerous occurrences of the eliciting stimulus. The ERP waveform is comprised of various "components," each having a characteristic scalp topography, latency range, and polarity. It is assumed that these components reflect the electrical activity from numerous generators within the brain, the activity of which overlaps in both space and time. For our purposes, it is not critical to understand the brain loci and generator mechanisms underlying these scalp-recorded components. Instead the focus is on how these components vary differentially with experimental manipulations and what these systematic variations suggest about the mental operations that these manipulations call into play. The components of most interest here are those which have been shown by previous work to be related, not to the physical characteristics of the stimuli to which an ERP was time-locked, but to the cognitive processing which was required by the task within which these stimuli were presented. Differential scalp topography and differential response to manipulations of the cognitive task are the primary means for disentangling the functional components of these waveforms.

Studies relating ERP components to mental workload grew out of previous findings which showed consistent attention-related effects on the amplitude of the P300 component. P300s are elicited by stimuli that are attended (i.e. task relevant) and, in some sense, unpredictable (see e.g., reviews in Donchin, et al., 1978). The basic hypothesis underlying most studies of P300 and workload has been that P300 amplitude would be modulated by the amount of attention, or the amount of central processing resources, that could be devoted to processing the ERP-eliciting stimuli. Thus, in dual-task situations, when the attentional demands of the primary task are increased, there is less of the limited pool of attention that can be devoted to secondary task stimuli, and hence the amplitude of the P300 elicited by such secondary task stimuli should decrease. Much of this work has been performed by Donchin, Wickens, and their colleagues, at the University of Illinois (see review in Donchin, et al., 1986). In the early studies, tracking a computer-driven cursor was used as the primary task. The secondary task involved the presentation of discrete stimuli which required either an overt response to which choice reaction time was measured, or a covert updating of a running count of the occurrence of some subset of the stimuli.

The initial results were somewhat discouraging. The amplitude of the P300 elicited by low probability auditory stimuli in a counting task was markedly reduced when the counting was performed concurrently with a visual-motor tracking task; however, there were no further systematic decreases in the P300 amplitude as the difficulty of the tracking task was increased, either by requiring that tracking be performed in two dimensions (Wickens, et al., 1977) or by increasing the bandwidth of the cursor in a one-dimensional tracking task (Isreal, Chesney, Wickens, & Donchin, 1980).

More encouraging results were obtained when the auditory counting task was time-shared with a visual monitoring task in which subjects detected directional changes in a simulated air traffic control display. In this situation, the P300 elicited by auditory stimuli decreased in amplitude as a function of the number of elements which subjects monitored (Isreal, Wickens, Chesney, & Donchin, 1980). The interpretation of these findings was consistent with the viewpoint which was emerging from behavioral studies at the time (e.g., Wickens, 1980) which posited that processing resources were segregated into multiple "pools." Thus P300 amplitude elicited by secondary task stimuli may have been modulated by the demands of the primary task when it involved visual monitoring, because the perceptual demands of these two tasks may have tapped the same pool of processing resources. On the other hand, the P300 amplitude elicited by secondary task auditory stimuli may not have reflected the workload dynamics of the tracking tasks, because the visual-motor demands of tracking tapped a different pool of resources.

Further evidence that P300 amplitude is related to available processing resources was sought by examining the reciprocity between the amplitudes of the P300 elicited in the context of primary versus secondary task stimuli in dual task paradigms. In order to elicit ERPs related to primary task processing, a task was developed which involved compensatory tracking with the cursor moving in discrete steps, rather than moving continuously as before. When subjects tracked these step changes in conjunction with a secondary task that consisted of counting occurrences of certain auditory stimuli, the amplitude of the P300 elicited by the secondary task stimuli decreased as the difficulty of the tracking task increased. However, when subjects were instructed to count occurrences of the cursor step changes in a given direction (i.e., the

secondary task stimuli were "embedded" in the primary task), the P300 elicited by the step changes increased in amplitude as the tracking task was made more difficult (Wickens, et al., 1983).

These studies provided valuable insights into the way in which cognitive resources are allocated in complex tasks. In addition, they established P300 amplitude as a sensitive index of the amount of processing resources, in a sense the degree of attention, that is devoted to particular classes of stimuli in complex tasks. However, possible practical applications of these results are subject to the previously discussed limitations of secondary task methodologies. Granted, the fact that measures of attention allocation can be extracted from ERPs elicited by stimuli being covertly counted, offers the possibility of applying a secondary task methodology without the need to burden the subject with additional manual response requirements (Donchin, et al., 1986). However, even when the stimuli being counted are embedded in the primary task, as was the case when subjects counted step changes in a cursor being tracked (Wickens, et al., 1983), the cognitive demands of the counting task are superfluous to the otherwise existent task demands. One question addressed by our empirical studies was to what extent ERPs elicited by stimuli in a single, complex task, as they are processed naturalistically, will reflect the cognitive workload demands of the situation.

1.4 Phase I Empirical Study

For our Phase I study, we designed a laboratory task which provided discrete stimuli to elicit ERPs and allowed for the manipulation of mental workload, but yet was analogous, in many ways, to the types of monitoring activities which are performed in operational environments. The richness of this task afforded the opportunity to relate the waveforms elicited by similar physical stimuli to a variety of information-processing constructs, but without requiring subjects to concentrate on more than one task at a time. Our interest was in determining the extent to which graded effects on ERP amplitude as a function of mental workload could be observed within the context of this single-task. Positive results will suggest the usefulness of ERPs as indicants of certain mental processes in any setting which offers the ability to time-lock recordings to a discrete eliciting stimulus, regardless of whether or not other

tasks are being performed concurrently.

1.4.1 Task Definition

The subject's task was to monitor successive CRT displays of a circular array of six two-digit readouts. On each presentation of the display, termed a trial, one of the six readouts changed from its value on the previous trial. The values of the readouts changed, either increasing or decreasing, in large (30) or small (10) steps, within the range from 00 to 99. Large step changes were less frequent than small step changes. Presentations of the array of readouts lasted 500 msec and were separated by intervals which varied randomly from 1800 to 1900 msec.

Subjects were instructed to monitor a subset of the readouts to determine which of these readouts reached 90 or above or fell to 10 or below. Readouts which met or exceeded these target values were referred to as having gone "out-of-bounds." Workload was manipulated by instructing subjects to monitor one (low workload), two (medium workload), or three (high workload) of the six readouts. After passively monitoring a "run" of twenty trials, subjects reported the positions and sequence of occurrence of targets, i.e. attended readouts that went out-of-bounds. A given subset of readouts was designated as the targets for a sequence of six successive runs. The order of these workload conditions and the arrangement of the target readouts were counterbalanced.

In the Phase I study, there was an equiprobable chance that each of the six readouts would change on a given trial. Thus the probability of a monitored readout changing was dependent on the number of readouts being monitored.

1.4.2 Rationale for the Task

In this present monitoring task, the way in which the stimuli varied from observation to observation was different from the method used in most studies in the literature. Typically, the sequence of stimuli in ERP studies consists of a Bernoulli series; i.e., the particular stimulus presented on each trial is independent of that presented on previous trials. The goal in designing the present experiments was to construct a monitoring task which called into play

the same cognitive processes that are invoked in real-world monitoring tasks. In operational settings, the likelihood of a particular meter reading or display state is determined by those of the recent past; drastic changes from the last reading are less likely than relatively small changes; readings which require an overt response, e.g. because they reflect a system with some parameters "out-of-bounds," are preceded by readings in the "danger" zone.

In reflecting these features, the monitoring task used was analogous to a wide variety of real-world challenges. A pilot's in-flight interaction with engine performance and environmental system displays or a process control operator's monitoring of plant status are fairly obvious examples of such circumstances. However, in terms of the cognitive processes invoked, the present task was also analogous, in perhaps less obvious ways, to other applied tasks. For example, an air traffic control display of planes moving about an airspace also presents information which, while not entirely predictable, is nevertheless dependent on trends. Monitoring such displays as planes move towards or away from "danger zones" and, at times, enter "out-of-bounds" conditions, such as impinging on another plane's circumscribed airspace, presents many of the same mental challenges as the present laboratory task.

This monitoring task afforded the opportunity to investigate a number of cognitive influences on ERPs. Selective attention effects on ERPs could be distinguished by comparing responses to changes in a readout being monitored as opposed to changes in a readout for which there was no such task requirement. Similarly, processing which specifically reflected the occurrence of a "target" stimulus, could be distinguished by comparing the responses elicited by attended readouts that went out-of-bounds to those elicited by attended readouts that stayed or went in-bounds, or those elicited by unattended readouts which changed in any manner. In addition, we were interested in the ERP effects related to both "tonic" changes in information processing workload, imposed by the number of readouts being monitored throughout a run of trials, and the more "phasic," dynamic influences imposed by the number of attended readouts that were close to, i.e. in "danger" of, going out-of-bounds.

It is interesting to consider how the pattern of effects related to these variables, aside from demonstrating the sensitivity of ERPs to these cognitive

influences, can reveal specific aspects of subjects' performance in the task. For example, the extent to which the ERPs reflect the influence of attention, the differences between targets and non-targets, or effects related to number of monitored readouts that are "in danger," might change with the level of "tonic" workload. Will the need to monitor more readouts cause a focusing of attention, and thus perhaps greater differences between responses to monitored and non-monitored readouts? Might increasing task demands cause target stimuli to be processed differently? Might the number of readouts "in danger" be more readily noticed when workload is high, because this information could be used by the subject to distinguish which of the readouts being monitored are most likely to become targets in the near future, or will this information be disregarded when workload is high, due to the fact that there are fewer central processing resources available to devote to this additional processing?

1.4.3 Phase I Results

The Phase I results indicated several ERP effects related to mental workload and attention (see Horst, et al., 1984; Horst, et al., 1985):

o   The amplitude of a long-latency, slow positivity in the average ERP waveforms elicited by both monitored and non-monitored readouts increased with the number of readouts being monitored

o   A slightly earlier region of the waveforms, which contained the peak positivity, increased in amplitude with the number of monitored readouts "in danger"; no such effect was found when the number of non-monitored readouts "in danger" were examined

o   This same peak positivity was markedly larger in response to "target" readouts, i.e., those that took an attended readout out-of-bounds, than in response to non-targets

o   The peak positivity was also somewhat influenced by the number of readouts being monitored; peak amplitude increased with such increases in workload, but surprisingly, this effect was seen only in responses elicited by the non-monitored readouts and not in the responses elicited by the monitored

readouts

o   No differences were found in the level of the pre-stimulus baseline region of these ERPs, suggesting that the workload effects influenced post-stimulus processing of the readouts.

The timing, polarity, scalp distribution and differences between the ERPs elicited by target and non-target readouts suggested that the peak positivity in these waveforms was due to the P300 component. The increasing amplitude of P300 with increases in workload were consistent with previous reports (e.g., Wickens, et al., 1983), although the lack of any effect in the responses to monitored readouts was perplexing. Because each of the six readouts changed with equal probability, as more readouts were monitored, the likelihood of a monitored readout changing increased, and the likelihood of a non-monitored readout changing decreased. Therefore, probability effects may have confounded a possible workload effect on the P300. The longer latency, slow positivity was likely related to the Slow Wave component (see e.g., Ruchkin & Sutton, 1983), although the scalp distribution for this component in the Phase I data was not characteristic of most Slow Waves reported in the literature. The finding of a workload-related effect on Slow Wave amplitude, in both the ERPs elicited by monitored and non-monitored readouts was unprecedented.

These results clearly suggested the feasibility of obtaining ERP indices of workload in a single-task paradigm that was analogous to monitoring tasks required in operational settings. However, questions remained as to the effects of "tonic workload" on P300, the nature of the slow positivity, and the generality of these results.

1.5   Overview of Phase II Study

The Phase II study was designed to address these issues and to examine the feasibility of transitioning ERP indices of mental workload from the laboratory towards operational settings. The purpose of such a transition would be to develop the use of ERP measures as tools for application in the systems engineering process. There may also be eventual uses of ERPs as real-time measures of operator performance for input to automated decision-making support

systems (see Horst, 1988).

The Advanced Concepts Flight Simulator (ACFS) at NASA Langley was chosen as a target environment for the first application of such ERP measures of workload. The ACFS represents a test-bed for the flight deck of a 1995 transport aircraft and has been developed jointly by NASA and Lockheed-Georgia Corporation for research purposes. As in many such complex man-machine systems, the determination of aircrew mental workload is a primary concern in making design decisions. It is hoped that psychophysiological measures such as ERPs, obtained from the pilot, may complement measures of workload based on behavioral or subjective data.

ARD's Phase II project involved five main tasks, which define a logical progression of the ERP measurement technology from a laboratory towards a simulator-based environment:

1) Two laboratory studies were conducted in order to explore, under controlled conditions, the generality of the encouraging ERP results obtained in the Phase I study.

2) A "task analysis" of flight scenarios and pilot decision-making in the ACFS was conducted for the purpose of defining events (i.e., displays, messages, alarms) to which the aircrew are exposed during realistic flight scenarios that would be expected to elicit ERPs related to workload.

3) Software was developed to support ERP data analysis; this task included three subtasks -- the upgrade of an existing ARD-proprietary package of ERP data analysis routines, the development of new routines for graphic displays to enhance interactive data analysis, and the development of routines to simulate single-trial ERP data for the purpose of systematically comparing two alternative single-trial analysis techniques.

4) Working in conjunction with NASA Langley research scientists and simulator engineers, preparations were made for a validation study of ERP measures of workload using the ACFS and laboratory facilities at Langley.

5) A design specification was developed for a general purpose, computerized, workload assessment system that can function in simulators such as the ACFS or in related operational environments.

For the present report, the single-trial analysis subtask will be dealt with separately from the other software development activities. The inter-relations of the various project tasks is shown in Figure 1-1. The preparations for a simulator-based validation of the present methodologies is seen to draw upon the ACFS "task analysis" as well as the results of the laboratory empirical studies. The design specification of a workload assessment system draws upon the results of the five other tasks and, in many respects, represents the culmination of the project as a whole.

In the following sections of the present report, each of these project tasks are reviewed in detail. For each task, background information is provided, methods are reviewed, results are summarized, and implications for the overall technical objectives of the project are highlighted.
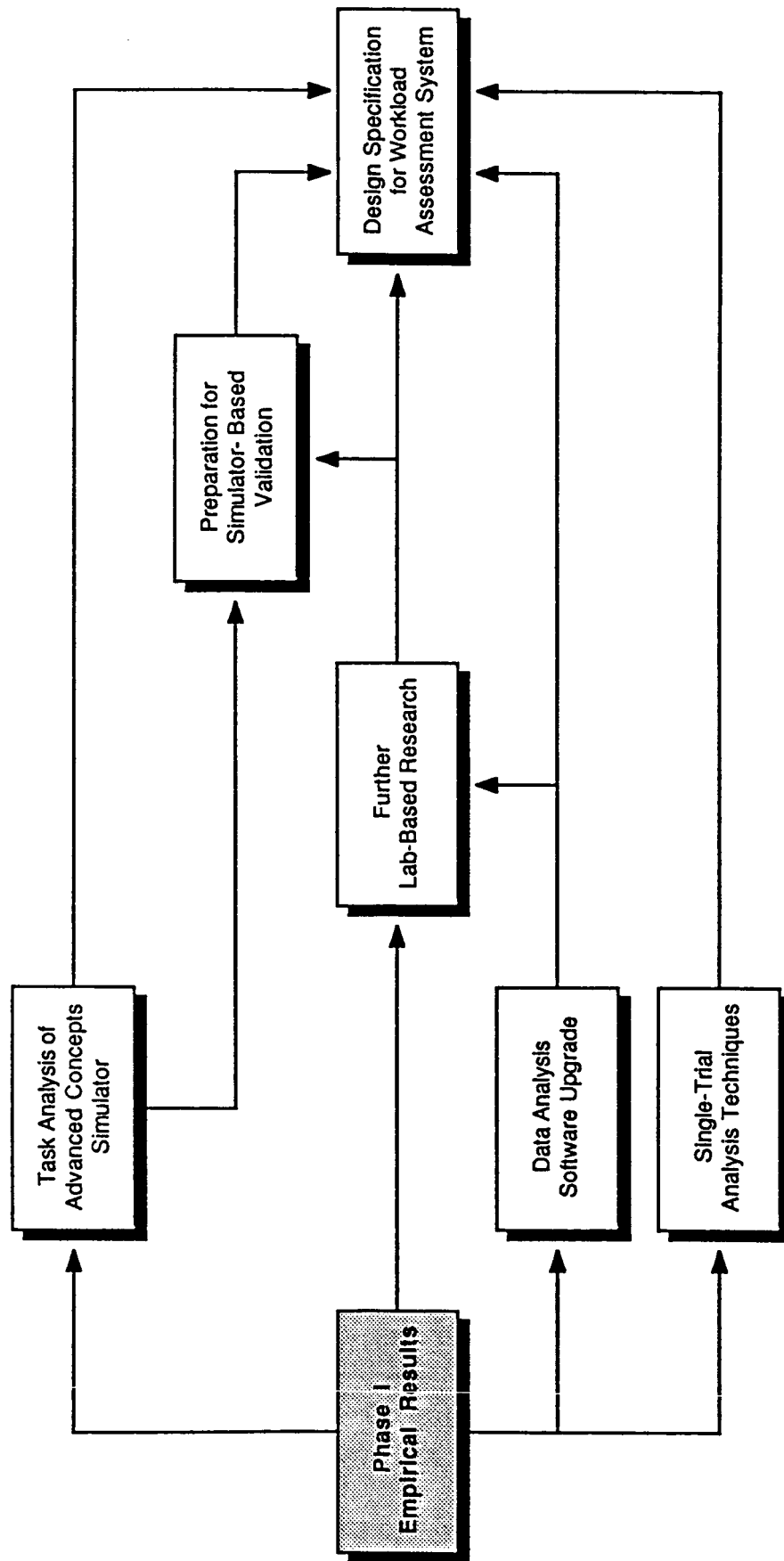
Figure 1-1. Overview of Project Tasks

2.0 FURTHER LABORATORY-BASED RESEARCH ON ERP INDICANTS OF MENTAL WORKLOAD

As discussed in Chapter 1, most previous investigations that have addressed the relationship between ERPs and mental workload have focused on responses elicited in dual-task paradigms. Typically the waveshape of the ERP elicited by secondary task stimuli has been related to changing levels of difficulty of the primary task and has been interpreted as reflecting the spare cognitive capacity that remains after the demands of the primary task have been met. While the results of these studies have revealed important insights regarding the influence of cognitive processes on ERPs, it is not clear how widely applicable this methodology will be in evaluating the workload of human operators in real-world systems.

The secondary tasks used in most laboratory studies of mental workload have been relatively simplistic and contrived. They have been chosen for the convenience with which their stimuli elicit the responses of interest, whether physiological or behavioral. Although such tasks offer a conceptual similarity to operational systems in which human operators must time-share between tasks and process stimuli which compete for attention, they do not lend themselves readily to use in operational or simulated systems. In most operational systems in which mental workload is a concern, the operator is already over-burdened. To further burden him with contrived stimuli and tasks, in order to assess the workload of existing tasks, is impractical at best and invalid at worst. Even if certain existing tasks offer stimuli to which ERPs, and reaction times, can be time-locked, it is unlikely that they will be functionally equivalent to the contrived secondary tasks used in the laboratory. Such "secondary" tasks will likely be performed in conjunction with differing configurations of other existing tasks, and it is difficult to ensure that these other "primary" tasks are given priority, as implicitly assumed if one is to interpret secondary task measures as reflecting spare cognitive capacity.

Because of these considerations, we examined ERPs that were elicited by stimuli presented in a single (primary) task as the difficulty of that task was varied.

Workload-related effects obtained in such a paradigm would suggest the usefulness of ERP measures of cognition, both for systems in which processing resources can be devoted to a single-task, as well as those in which the ERP-eliciting task must be time-shared with others.

2.1 Objectives of the Phase II Empirical Studies

Our previous work has demonstrated differences in cognitive-related ERP responses in complex tasks. In our previous investigation, subjects monitored a circular display of six two-digit readouts. On each trial, the value of one of the readouts changed. Only a subset of the display was salient to subjects in that they were instructed to monitor one, two, or three of the readouts and to report values in these monitored readouts that exceeded pre-specified boundaries. There were larger amplitude late positivities at approximately 600 msec post-stimulus onset on trials where a change occurred in a salient (i.e., monitored) as opposed to non-salient (i.e., non-monitored) readout. The amplitude in this region also increased with increasing workload, but only for responses to the non-monitored readout changes. Responses to changes in monitored readouts showed no apparent effect of workload. Consequently, the difference between monitored and non-monitored readouts decreased as the number of monitored readouts increased. However, there were workload-related amplitude differences at approximately 1000 msec after stimulus onset. ERP amplitude in this region increased as the number of readouts being monitored increased, regardless of whether the change occurred in a monitored or non-monitored readout. The results of this study indicated that, under some circumstances, different portions of the ERP waveform can reflect attention and workload differences in complex tasks.

In this previous study, it was equiprobable that each readout would be the one to change on a given trial. Therefore, as the subject monitored either one, two, or three readouts, the probability of a change in a monitored (vs. non-monitored) readout changed. It was possible that these varying probabilities contributed to the differences seen in the effects of workload on responses to changes in monitored and non-monitored readouts. Thus, in the first of the current experiments, the probabilities of a monitored and non-monitored readout changing were made equal, regardless of how many readouts

were being monitored. In Experiment 2, we examined the extent to which the obtained ERP effects related to attention and workload generalized to a situation in which subjects responded trial-by-trial, rather than holding in memory the information about how many and which attended readouts had gone out-of-bounds.

2.2 Methods

2.2.1 Subjects

In each of the two experiments, 12 young adult males served as subjects. Subjects were paid $7.00 per hour for their participation. All subjects had normal or corrected-to-normal vision.

2.2.2 Apparatus

Both experiments were programmed on and controlled by a DEC PDP 11/40 minicomputer. The computer controlled stimulus delivery, behavioral and physiological data acquisition, and data storage. Subjects were seated in a darkened, electrically shielded room approximately two and one-half feet from the CRT display monitor. Each display subtended a visual angle of approximately 1.8 by 2.3 degrees of visual angle. ERPs were recorded using silver/silver chloride electrodes from five mid-line sites (Fpz, Fz, Cz, Pz, and Oz). In addition, EOG was recorded from electrode placements above the inner and below the outer canthi of the left eye.

2.2.3 General Characteristics of the Task

The aspects of the two experimental tasks that were the same for both studies are described here. Task elements specific to each experiment are described later. The subject's task in both experiments was to monitor successive CRT presentations of a circular display of six two-digit numbers. The display was arranged in a clock-like pattern with the two-digit "readouts" at the 2, 4, 6, 8, 10, and 12 o'clock positions. Subjects were instructed to monitor a prespecified subset of the readouts. On each trial (i.e., presentation of the display) one of the six readouts changed from its value on the previous trial.

Changes in a readout being monitored occurred equally often as changes in a non-monitored readout.

Subjects monitored the display to determine which of the monitored readouts reached 90 or above or fell to 10 or below. Readouts which met or exceeded these target values were referred to as having gone "out-of-bounds." Workload was manipulated by instructing subjects to monitor one (low workload), two (medium workload), or three (high workload) of the six readouts. The order of the workload conditions and the arrangement of the target readouts were counterbalanced.

Stimuli were presented in a sequence of 20 discrete displays called "trials." The subject initiated a sequence of 20 trials (a "run") by actuating an optical switch. Presentations of trials lasted 500 msec and were separated by intervals which varied randomly from 1800 to 1900 msec. For a "set" of six consecutive runs, subjects monitored the same subset of readouts. At the beginning of each set, all readouts were between 40 and 60. Within a set, readout values were maintained across runs, except those readouts that were out-of-bounds at the end of a run. These out-of-bounds readouts were reset to a randomly chosen in-bounds value. Throughout, readouts that would have exceeded 99 were presented as 99, and readouts that would have gone below 0 were presented as 00. At the beginning of a new "set" of trials, the first display presented to the subject read "new," indicating that a different subset of the display was to be attended. Before each run of trials, a display appeared to remind the subject which readouts were to be monitored. This display was a circular, clock-like pattern (like the "run-time" display) with the numbers one, two, and/or three in the readout positions to be monitored. The appearance of this display informed subjects that they could begin the run by actuating the optical switch. There was a total of 2160 trials (18 sets; 108 runs) in each experiment. During a run, subjects were asked to focus their gaze on a centrally located fixation point, rather than looking directly at any one of the readouts.

In both experiments, subjects were given 15 to 30 minutes of practice, or they were trained until it was clear that they understood the instructions and were responding appropriately.

2.2.4 Task Characteristics Specific to Experiment 1

On each trial (i.e., presentation of the circular display of readouts), one of the six readouts changed from its value on the previous trial. The values of the readouts changed, either increasing or decreasing, in large (30) or small (10) steps, within the range from 00 to 99. Large step changes were less frequent (33%) than small step changes (67%). The direction of the change was completely random and could go in either direction, regardless of the direction of the last change.

After passively monitoring a "run" of 20 trials, subjects reported the positions and sequence of occurrence of targets, i.e. attended readouts that went out-of-bounds. This was done by pressing buttons on a response box resting in the lap of the subject. It was important for the subject to report both the order and the location of the positions that went out-of-bounds. If, for example, readout 2 went out-of-bounds, then readout 1 went out-of-bounds, then (after coming back in-bounds) readout 2 went out-of-bounds again, the correct response, for that run, would be "2-1-2." Subjects were given run-by-run feedback as to their accuracy as well as the correct sequence of responses.

In a separate "behavioral" data collection session, each subject was presented with all of the conditions under which ERP data had been collected. On each trial, subjects indicated, with a two-choice reaction time response, whether or not a monitored readout had gone out-of-bounds.

2.2.5 Task Characteristics Specific to Experiment 2

In Experiment 2, as in Experiment 1, only one readout changed per trial; however, unlike the first experiment, the readout always changed by a fixed value of 20. Another difference in Experiment 2 was that the direction of the change (increasing or decreasing) was not equiprobable. There were, in fact, three levels of "trend" determining the direction of change for the next readout. The trend levels were: high — 90% of the time a change in value for a given readout would continue in the same direction as the previous change for that readout; medium — 70% of the time the change for a given readout would

continue in the same direction as the previous change for that readout; and low — 50% of the time a change for a given readout would continue in the same direction as the previous change for that readout. The low (50%) level of trend is equivalent to the directional probabilities in Experiment 1.

In Experiment 2, subjects reported after each trial as to whether or not a monitored readout had gone out-of-bounds. There were two alternative ways in which subjects made their reports. The first response condition was called "go on targets." In this condition, when a monitored (or target) readout went out-of- bounds, the subject made a small finger movement with the index finger of the right hand to actuate an optical switch. Otherwise, the subject inhibited this response. The second response condition was called "no-go on targets." In this condition, subjects made the finger movement on all trials except when a monitored readout went out-of-bounds. Thus, in this second response format, when a monitored readout went out-of-bounds, the subject did not make the finger movement he was otherwise making.

ERP data were collected under several additional "control" conditions in Experiment 2. These control conditions were as follows:

o   A display of the same six readout values was repetitively presented for 50 trials, with subjects being instructed to passively view it with their eyes fixated on a particular readout. The purpose of this condition was to determine the waveshape of the sensory ERP elicited when subjects looked directly at a readout. It was hoped that the sensory-related early components in these waveforms would differ from those seen in the waveforms collected during the monitoring task conditions, thus supporting the assumption that, in the latter conditions, subjects successfully followed the instructions to maintain eye fixation on the centrally located fixation point during the task conditions.

o   A display of the same six readout values was repetitively presented for 50 trials, with subjects being instructed to respond to each as quickly as possible with a finger movement (i.e., simple reaction time). The purpose of this control condition was to determine the ERP waveshape elicited by the readout stimulus presentation and motor response, but without the

cognitive decision-making that was required during the monitoring task.

o   With no display being presented on the CRT screen, subjects made 50
    self-paced finger movements of the sort required in responding during the
    monitoring task conditions. The purpose of this control conditions was to
    determine the waveshape of the average motor potential that could be
    assumed to affect the stimulus-locked ERP waveshapes when subjects
    responded to the readout displays with a finger movement (i.e., readout
    changes which took a monitored readout "out-of-bounds" during the "go on
    target" condition and "in bounds" readout changes during the "no-go on
    target" condition.

2.2.6   Processing of ERPs

The ERP waveforms from trials on which the EOG channel indicated that a blink
or eye movement had occurred were rejected from further analysis. Likewise,
the ERPs from trials on which an incorrect behavioral response occurred were
rejected.   Therefore, average ERPs were formed only from trials on which a
correct response and no substantial EOG activity had occurred. The waveform
data were sorted according to the various condition variables of interest and
the resulting average ERPs were quantified as indicated below. ALL OF THE
EFFECTS DISCUSSED HERE WERE STATISTICALLY SIGNIFICANT AT THE P < .05 LEVEL OR
BETTER, AS INDICATED BY MULTIVARIATE ANALYSIS OF VARIANCE.

2.3   Results of Experiment 1

Both behavioral and ERP results are discussed below. The ERP effects are
distinguished as follows:

o   "Target effects" -- effects related to whether or not the ERP-eliciting
    readout change took a readout being monitored out-of-bounds.

o   "Selective attention effects" -- effects related to whether or not the
    ERP-eliciting readout change occurred in a monitored or non-monitored
    readout.

o   "Tonic workload effects" -- effects related to whether or not the ERP-eliciting readout change occurred in a run of trials during which workload was low, i.e., only one readout was being monitored, versus a run during which workload was high, i.e., three readouts were being monitored.

o   "Phasic effects" -- effects related to the number of monitored readouts that were "in danger" of going out-of-bounds, i.e. those that were within one large jump of going out-of-bounds when the ERP-eliciting readout change occurred.

2.3.1   Behavioral Results

Figure 2-1 presents the reaction time data from the "behavioral" test session. Analysis of variance indicated that the trend towards increasing reaction times with increasing number of readouts being monitored was highly significant. The results provide converging evidence that by requiring subject to monitor differing numbers of readouts, we were, indeed, manipulating workload.

2.3.2   General Aspects of Obtained ERPs

There were several aspects of the averaged ERP waveforms obtained here which showed systematic variations in response to one or more of the factors of interest. These features were designated and quantified as follows:

1)   The "peak positivity" -- the mean amplitude over a 200 msec epoch centered about the most positive peak between 500 and 900 msec post-stimulus onset.

2)   The "slow positivity" -- the mean amplitude between 900 and 1050 msec post-stimulus onset.

3)   The N250 -- the mean amplitude between 200 and 300 msec post-stimulus onset.

4)   The N450 -- the mean amplitude between 400-500 msec post-stimulus onset.

Although ERP waveshapes were generally similar across subjects, there was
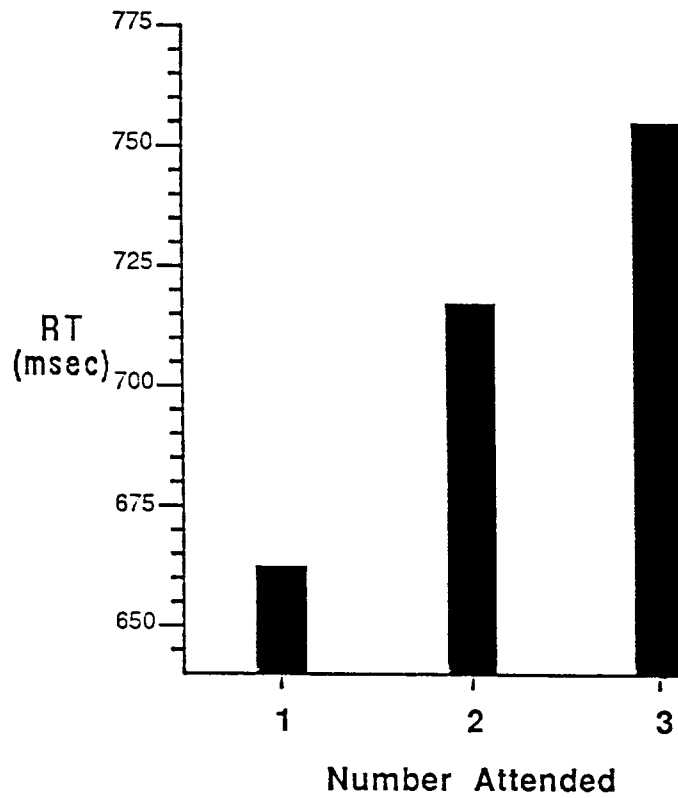
Figure 2-1. Mean reaction times, averaged across subjects from behavioral runs in Experiment 1.

considerable inter-subject variability in the latency of the peak positivity. These measurement epochs were selected after inspection of across-subject, grand-average waveforms, and were chosen to accommodate the systematic differences in the waveforms despite this latency variability.

Figure 2-2 presents across-subject, grand-average waveforms from Experiment 1, obtained from the Cz electrode. Responses elicited under low and high workload are superimposed. Responses elicited under the medium level of workload were intermediate to those shown here. In the various columns are responses elicited by readouts that were attended or unattended and which took the readout out-of-bounds, as opposed to taking or leaving it in-bounds. The waveforms elicited by target stimuli, i.e., monitored readouts that went out-of-bounds, are presented in the right-most column. Figure 2-3 presents difference waveforms in a similar layout to that of Figure 2-2. These difference waveforms were calculated by subtracting, for each condition and each subject, the waveforms elicited under low workload from the corresponding waveforms elicited under high workload. The resulting difference waveforms were then grand-averaged across subjects, to produce the plotted waveforms.

2.3.3 Target Effects

The most striking differences in the raw average waveforms of Figure 2-2 were the large increases in the amplitude of the late positivity for the responses elicited by target stimuli, i.e. monitored readouts that went out-of-bounds. This effect was limited to the region of the peak positivity and probably reflects a modulation of P300 amplitude that has been reported numerous times in the past (e.g., Duncan-Johnson & Donchin, 1977).

There was an additional target effect, this one related to tonic workload, that was evident in the difference waveforms. Figure 2-3 shows a negative-going wave in the 400-500 msec latency region that was present only when the responses to target stimuli elicited under low workload were subtracted from the responses to target stimuli elicited under high workload. Whether this waveform component should be seen as a negativity that enters in as the result of increased workload or a positivity that enters in as workload is reduced, can not be resolved.

Raw Average Waveforms
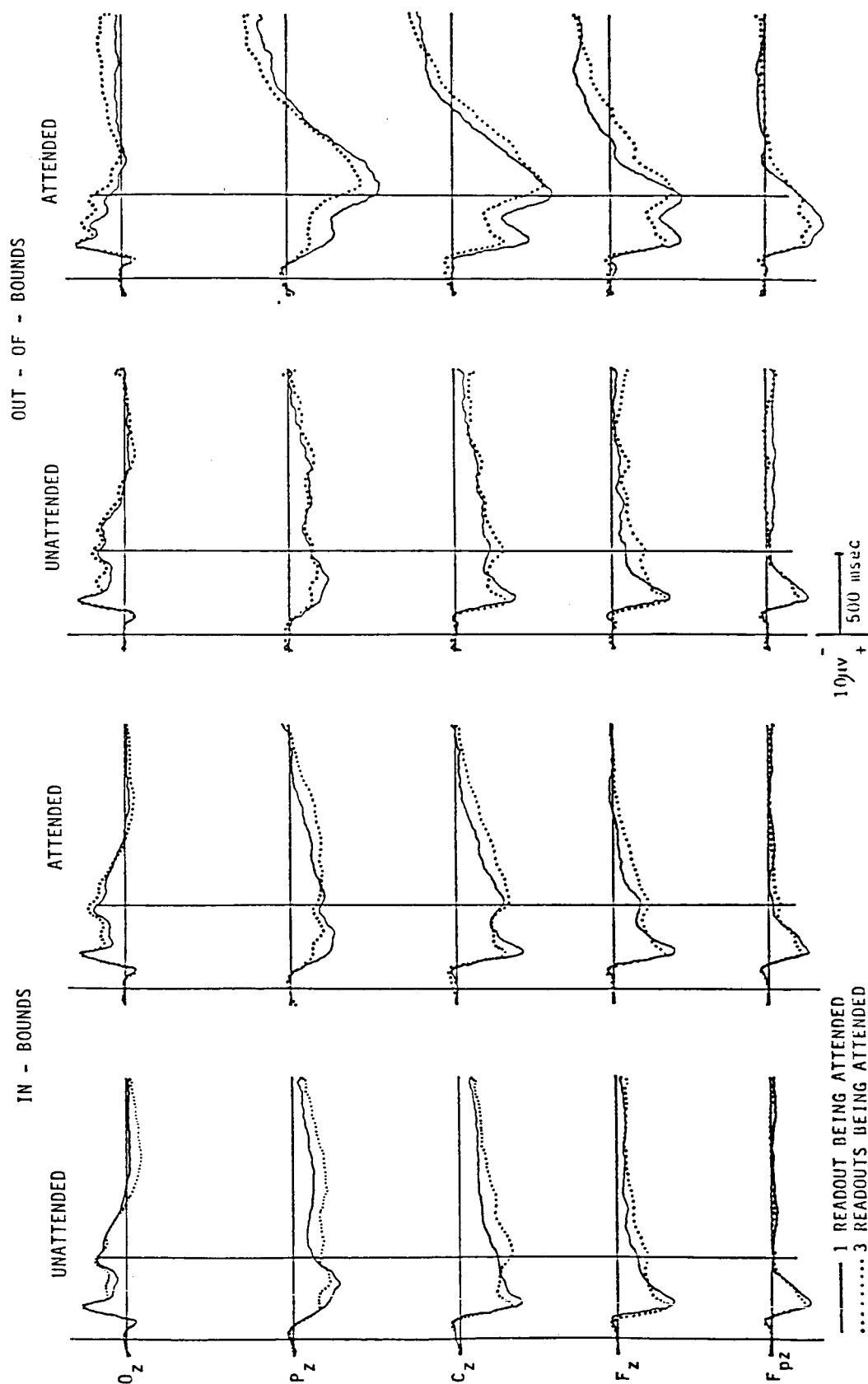
Figure 2-2. Across-subject average waveforms from Experiment 1, illustrating the effects of attention, target value, and tonic workload.

Difference Waveforms
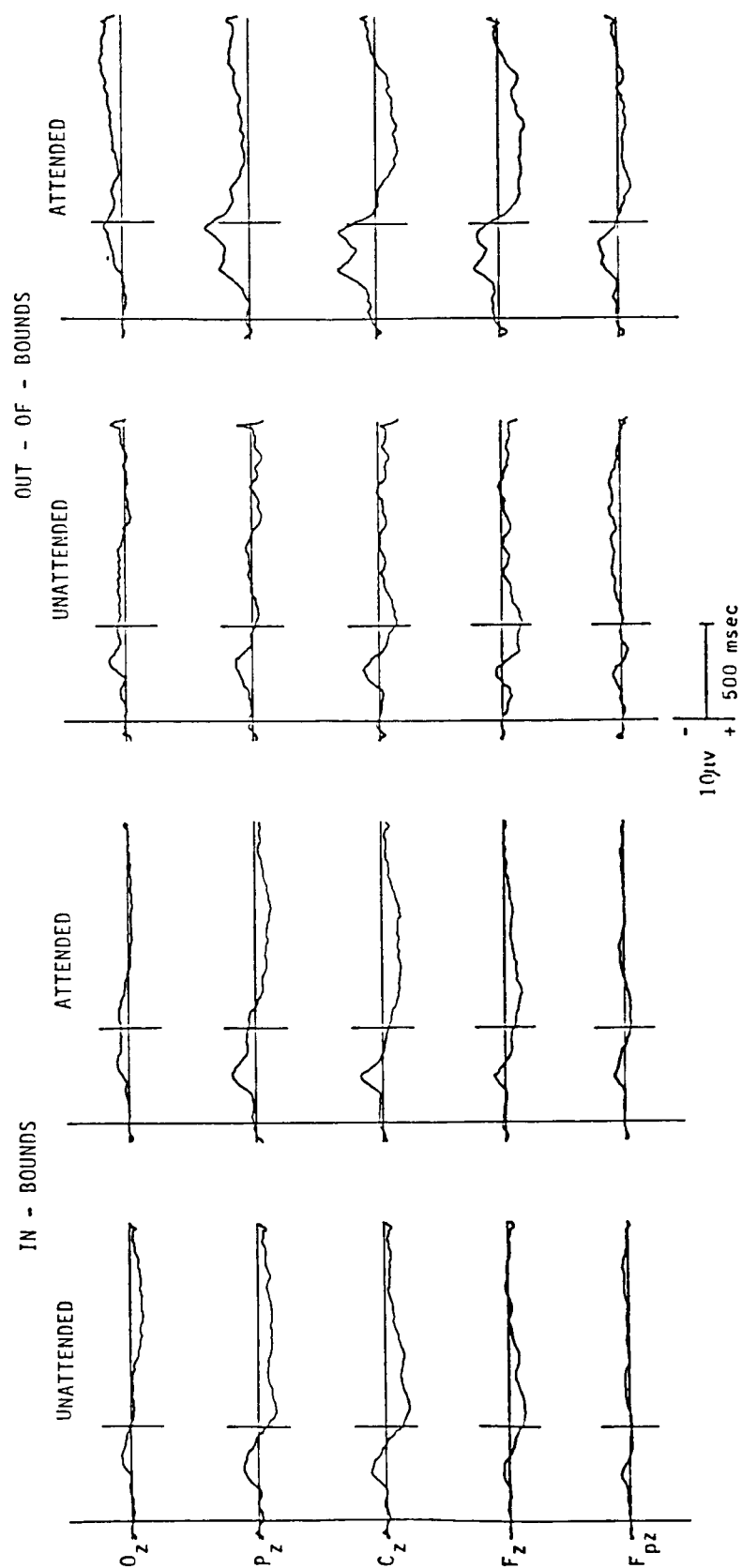
Figure 2-3. Across-subject average difference waveforms illustrating the effects of tonic workload.

However, the present results provide strong evidence that the workload manipulation added or enhanced a _new_ component in the waveform, rather than simply modulating a peak, or peaks, that were otherwise there. Peaks in a difference waveform that are due to either increases or decreases in amplitude, or to shifts in latency, of peaks that were evident in the raw average waveforms, should have the same scalp distributions as those raw average peaks. Instead, a comparison of Figures 2-2 and 2-3 indicates that the ERP peak in the 400-500 msec region of the difference waveforms had a more posterior distribution than either of the peaks in this vicinity of the raw average waveforms. This impression was confirmed by statistically showing that the profile of amplitudes across the scalp in this time region was different for the raw average waveforms elicited under low workload than for those elicited under high workload. Past references to endogenous ERP negativities in this latency region (e.g., Ritter, Simson, & Vaughan, 1983) provide a preliminary basis for interpreting this effect as an N450 component that is enhanced as the result of increased workload.

2.3.4 Selective Attention Effects

As can be seen in Figures 2-2, there was, at least at the low workload levels, a systematic difference between the ERPs elicited by changes in monitored (i.e. attended) and non-monitored (i.e. unattended) readouts. The amplitude of the peak positivity was larger in response to changes in monitored readouts as compared to changes in non-monitored readouts. This difference is best seen by comparing the responses elicited by in-bounds changes in the monitored and non-monitored readouts. Interestingly, the attention-related effect diminished with increasing workload, apparently due more to increasing peak positivities in the responses elicited by non-monitored readouts than to those elicited by monitored readouts. This same pattern of results was found in our Phase I data, when probabilities varied with the number of readouts being monitored. The differences between ERPs elicited by monitored and non-monitored readouts at low workload may be related to selective attention differences that have been interpreted as reflecting the activation of different sensory channels (e.g., Harter & Aine, 1984, Mangun & Hillyard, 1987); however, the polarity and timing of this effect, and its modulation by workload, is difficult to interpret. Further investigation of this effect is needed.

2.3.5 Tonic Workload Effects

Of primary concern in these data was whether there were differences in the ERP as a function of the level of workload imposed by requiring subjects to monitor different numbers of readouts. Two interactions with workload have already been noted — with increasing workload, an N450 component emerged in the responses to target stimuli and the peak positivity increased in the responses to all changes in non-monitored readouts. In addition, two main effects of the tonic workload manipulation are evident in Figure 2-2 and 2-3. First, as the subject was required to monitor an increasing number of readouts, the ERPs elicited by all stimuli showed an increased slow positivity. This slow positivity was manifest in the latency region following the peak positivity and can be seen as a slow return to baseline, but with a more posterior scalp distribution than the peak positivity itself. It is likely, although not entirely clear, that this slow positivity is the Slow Wave component which has been distinguished from the P300 on the basis of both scalp distribution and relationship to experimental manipulations (e.g., Ruchkin & Sutton, 1983).

A second main effect of tonic workload was apparent in the difference waveforms. When responses to readout changes from the low workload condition were subtracted from the corresponding responses from the high workload condition (Figure 2-3), a negative-going peak appeared in the 200-300 msec latency region. This N250 occurred in the responses to both changes in monitored and non-monitored readouts, regardless of whether these changes took the readout out-of-bounds or took or left it in-bounds. As with the N450, which was only present in the responses to target stimuli, we interpreted this N250 as a negative-going component which entered or was enhanced as the result of increasing workload. This interpretation was based on the fact that the scalp distribution of this wave differed from that of the corresponding activity in the raw average waveforms and the fact that processing negativities related to selective attention have been reported in this latency region of ERP waveforms (see e.g., Naataanen, 1982). Statistical tests confirmed that the amplitude profile across the scalp in the 200-300 msec latency region differed between the low and high workload conditions. To our knowledge, this workload-related effect had not been reported previous to our data.

It is possible that the standing requirement to monitor a given number of readouts for minutes at a time may have caused differential DC-shifts in the EEG. The transient ERPs elicited by readout changes might then have been superimposed on different baselines, and the apparent main effects of workload on post-stimulus ERP components could have resulted from a confound of, or interaction with, such differential baselines. To determine whether or not such differential pre-stimulus activity could have influenced the present findings, we did the recordings for Experiment 1 in a manner which allowed us to quantify the DC level of the pre-stimulus baselines. There were no systematic differences in the pre-stimulus baselines of the ERPs elicited under different workload conditions.

2.3.6 Phasic Effects of the Number of Readouts in Danger

As mentioned previously, the specific value of the readout presented on a given trial was dependent on its value on the previous trials; namely, it increased or decreased by a large or small increment from its value on the previous trial. Therefore, at any given time, only those readouts that were within a large increment of going out-of-bounds were "in danger" of becoming targets on the next presentation. Although it was not part of the subject's defined task to attend to this aspect of the situation, and no mention was made of it in the instructions, subjects could have facilitated their performance on the task by attending to this information. Therefore, we sorted the ERPs that were elicited with different numbers of readouts "in danger," to see if the waveforms showed evidence of this factor having influenced the processing of the readouts.

Figure 2-4 presents the data sorted for this analysis, with the responses superimposed that were elicited when 0, 1 or 2 monitored readouts were "in danger." These waveforms showed an enhanced positivity in the long latency regions with increasing numbers of monitored readouts in danger. Statistical tests confirmed this effect for the peak positivity, with the slow positivity showing the same trend but not reaching statistical significance. This increased positivity was present in both the responses to monitored and non-monitored readouts and was found to the same extent at all levels of tonic workload. When the waveforms were sorted as to the number of non-monitored

Across-subject average waveforms from Experiment 1 at Pz, Cz, and Fz, for trials on which a change in the display did not take the readout out-of-bounds. The superimposed waveforms are from responses elicited when different numbers of attended readouts were "in danger," i.e., within an incremental value of going "out-of-bounds." The left column contains trials on which a small jump occurred, i.e., a change of 10 in the value of the attended readout, and the right column contains trials on which a large jump occurred, i.e., a change of 30 in the value of the attended readout. The responses are sorted according to whether the eliciting change occurred in an attended or an unattended readout, and the number of readouts being attended at the time.

2-4-A   Change in an attended readout, when one readout is being attended

2-4-B   Change in an attended readout, when two readouts are being attended

2-4-C   Change in an attended readout, when three readouts are being attended

2-4-D   Change in an unattended readout, when one readout is being attended

2-4-E   Change in an unattended readout, when two readouts are being attended

2-4-F   Change in an unattended readout, when three readouts are being attended

EXPERIMENT 1 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-4-A. Change in an attended readout, when one readout is being attended
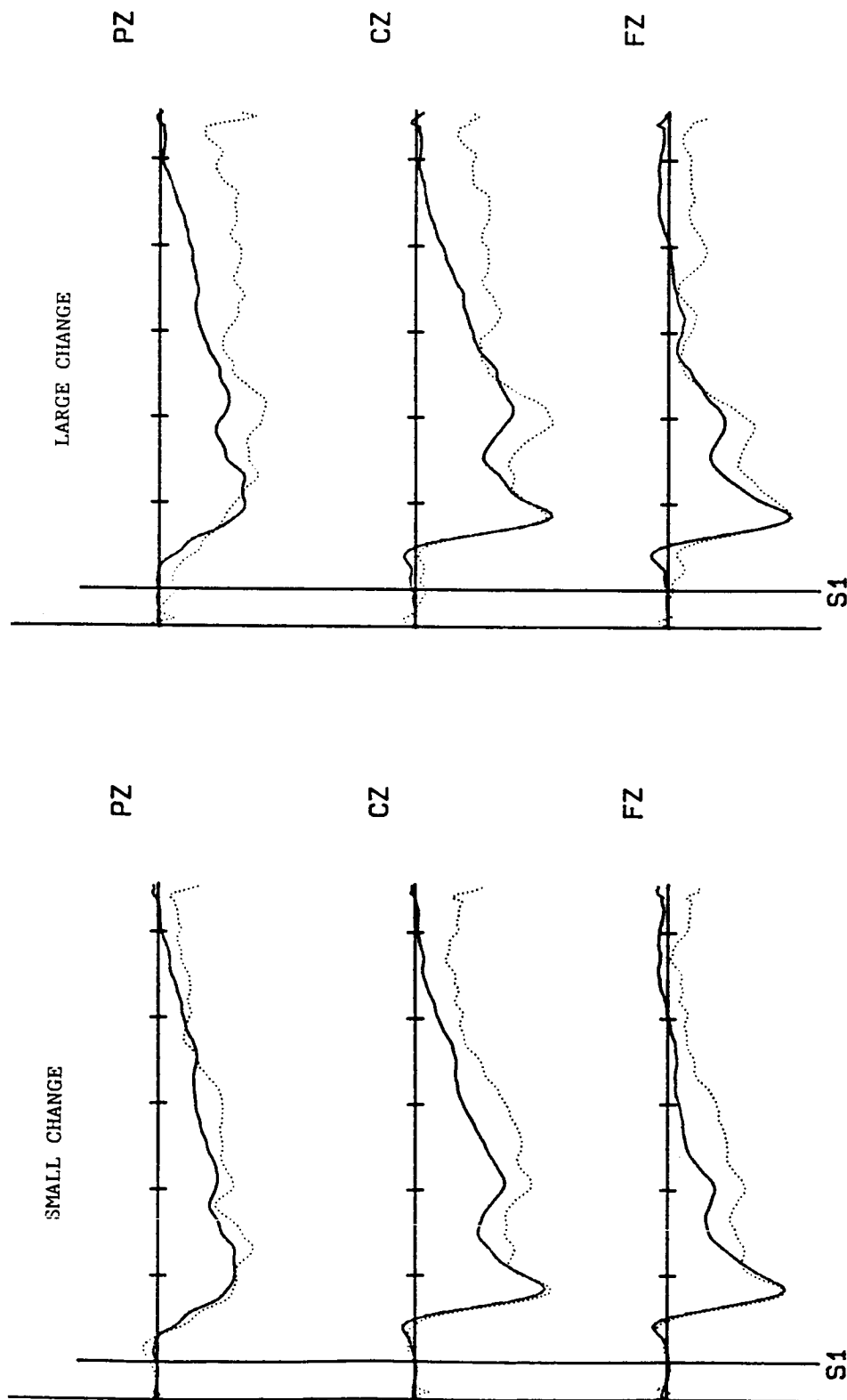
SMALL CHANGE

LARGE CHANGE

PZ

CZ

FZ

S1

15μV

450ms

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT
WHICH KEPT THAT READOUT IN-BOUNDS,
WHEN LOW WORKLOAD (ONE READOUT BEING MONITORED),
AND WHEN:

——— 2 MONITORED READOUTS IN DANGER

·········· 1 MONITORED READOUTS IN DANGER

———— 0 MONITORED READOUTS IN DANGER
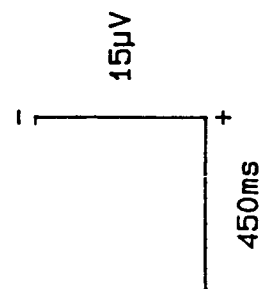
EXPERIMENT 1 — RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS
Figure 2-4-B. Change in an attended readout, when two readouts are being attended
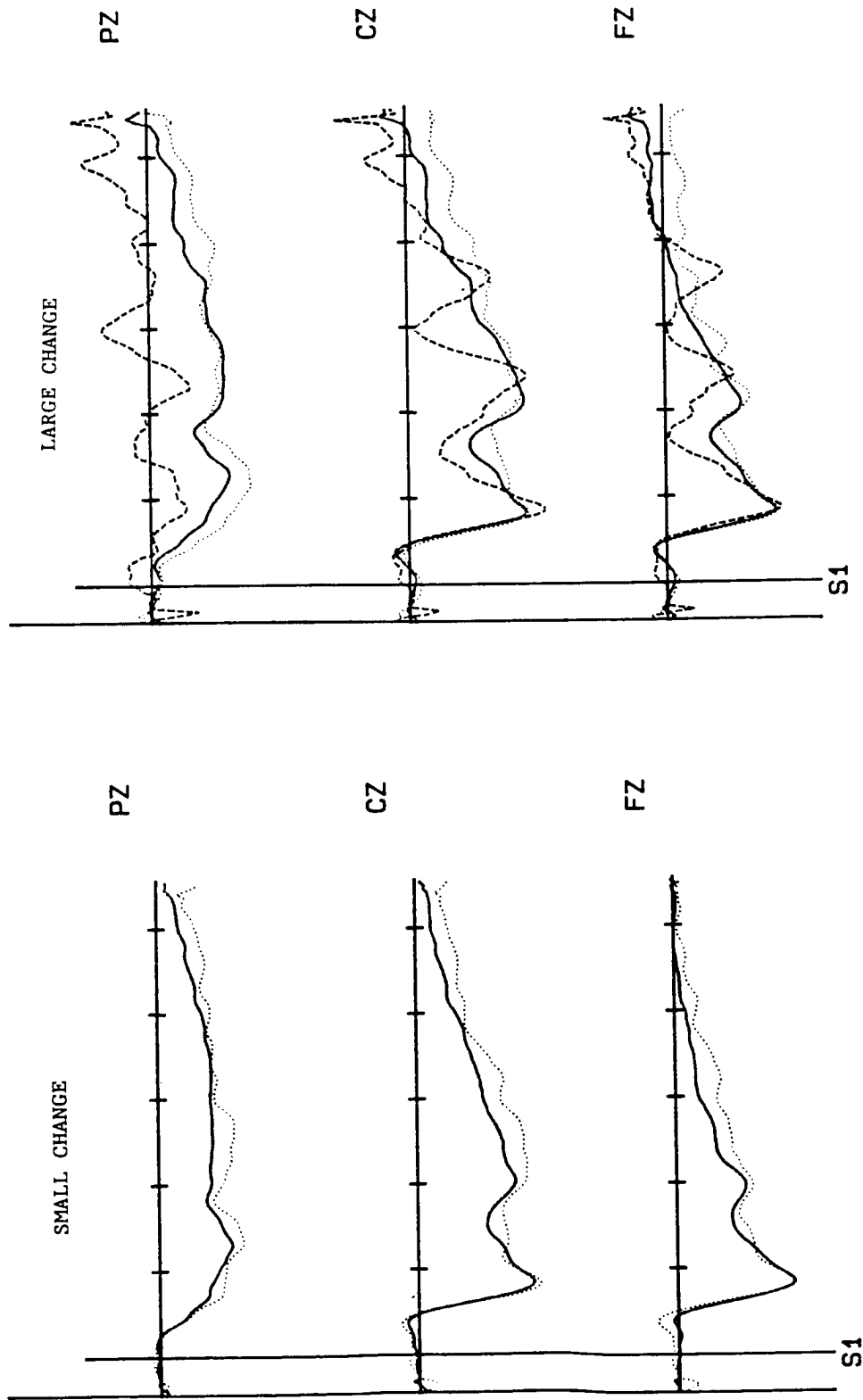
SMALL CHANGE

PZ

CZ

FZ

S1

LARGE CHANGE

PZ

CZ

FZ

S1

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT
WHICH KEPT THAT READOUT IN-BOUNDS,
WHEN MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED),
AND WHEN:

——————— 2 MONITORED READOUTS IN DANGER

············ 1 MONITORED READOUTS IN DANGER

——————— 0 MONITORED READOUTS IN DANGER

15μV

−

+

450ms

EXPERIMENT 1 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-4-C. Change in an attended readout, when three readouts are being attended

SMALL CHANGE

LARGE CHANGE

PZ

CZ

FZ

S1

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT
WHICH KEPT THAT READOUT IN-BOUNDS,
WHEN HIGH WORKLOAD (THREE READOUTS BEING MONITORED),
AND WHEN:

- - - - - - - - 2 MONITORED READOUTS IN DANGER

· · · · · · · · 1 MONITORED READOUTS IN DANGER

- · - · - · - 0 MONITORED READOUTS IN DANGER

15µV

450ms
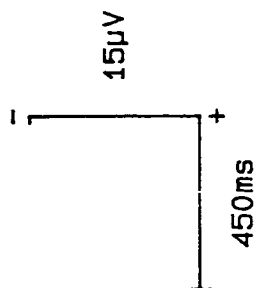
EXPERIMENT 1 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-4-D. Change in an unattended readout, when one readout is being attended

ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT WHICH KEPT THAT READOUT IN-BOUNDS, WHEN LOW WORKLOAD (ONE READOUT BEING MONITORED), AND WHEN:

- - - - - - - 2 MONITORED READOUTS IN DANGER

· · · · · · · 1 MONITORED READOUTS IN DANGER

- · - · - · 0 MONITORED READOUTS IN DANGER

EXPERIMENT 1 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-4-E. Change in an unattended readout, when two readouts are being attended
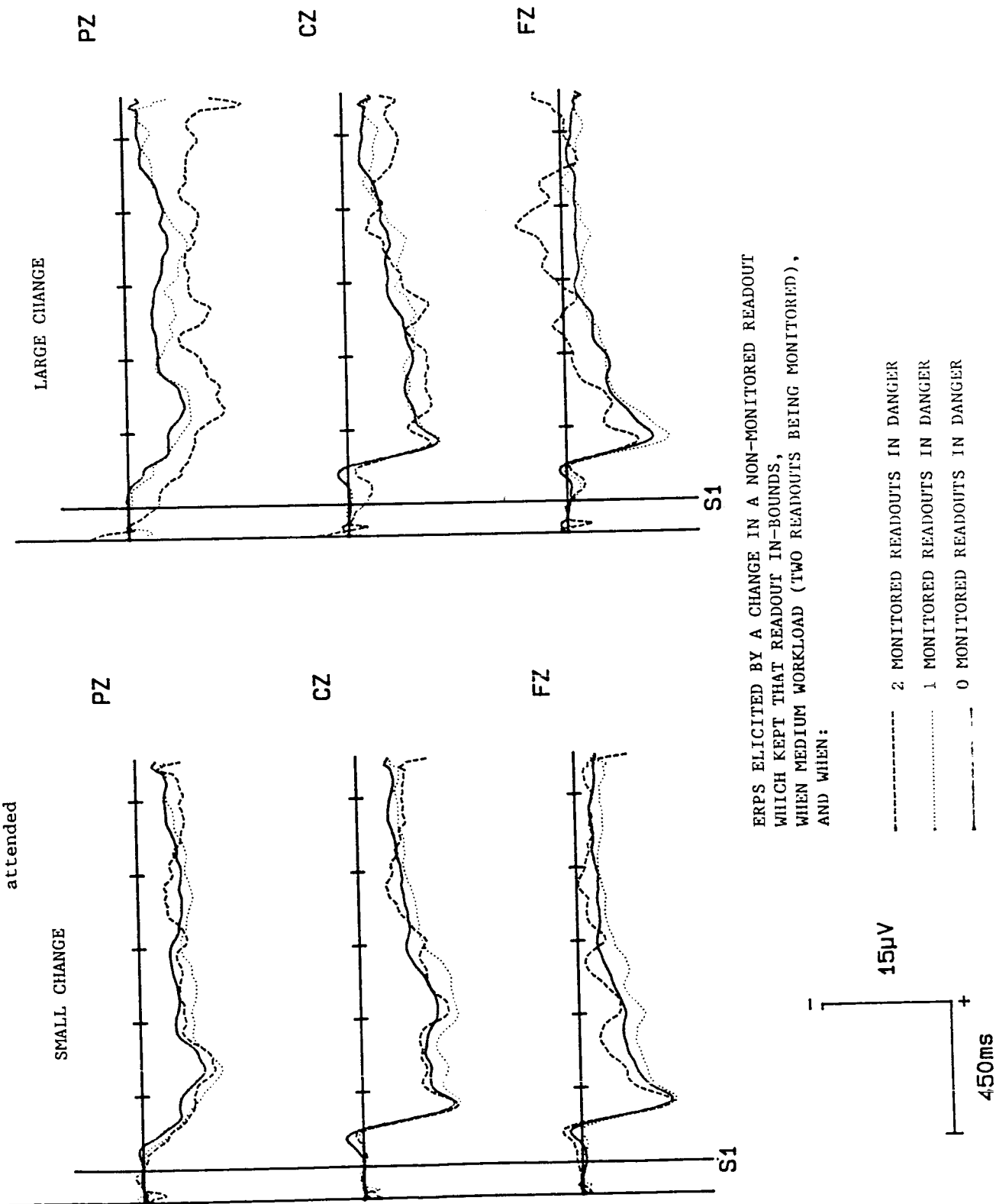
LARGE CHANGE

SMALL CHANGE

PZ

CZ

FZ

S1

15μV

450ms

ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT
WHICH KEPT THAT READOUT IN-BOUNDS,
WHEN MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED),
AND WHEN:

———————— 2 MONITORED READOUTS IN DANGER

················ 1 MONITORED READOUTS IN DANGER

—·—·—·— 0 MONITORED READOUTS IN DANGER

EXPERIMENT 1 — RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-4-F. Change in an unattended readout, when three readouts are being attended
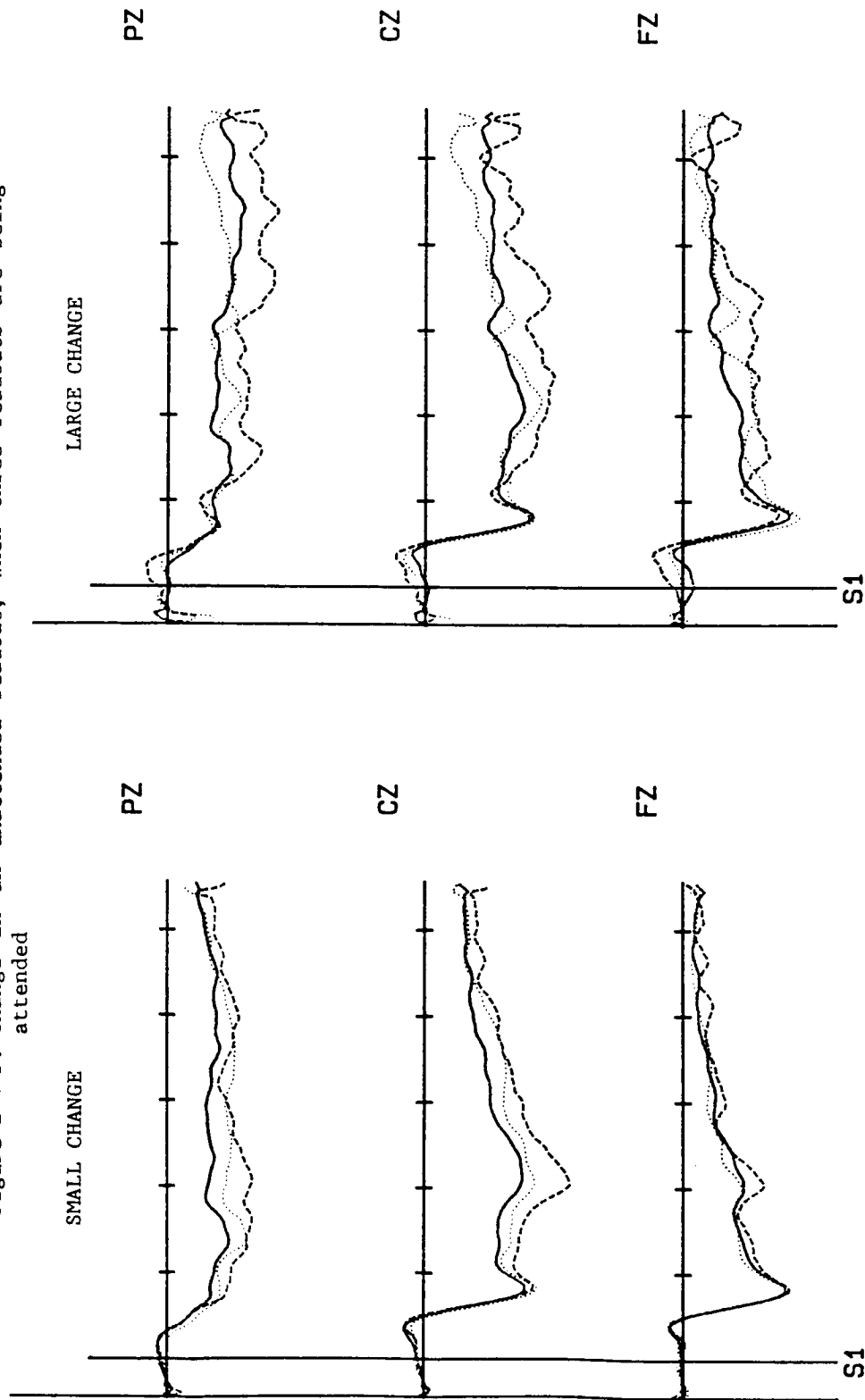
ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT WHICH KEPT THAT READOUT IN-BOUNDS, WHEN HIGH WORKLOAD (THREE READOUTS BEING MONITORED), AND WHEN:

——————— 2 MONITORED READOUTS IN DANGER

·············· 1 MONITORED READOUTS IN DANGER

— · — · — 0 MONITORED READOUTS IN DANGER

readouts "in danger," no systematic ERP differences were found.

These data, which also replicate our Phase I findings, clearly suggest that subjects processed the readouts differently depending on the number of monitored readouts that were close to going out-of-bounds, even though they were not explicitly instructed to do so. It is not clear whether this differential processing should be seen as an additional, albeit self-imposed, workload demand of the task, or whether subjects chose to assume this additional processing as a means of coping with the primary task of detecting target readouts. A number of further manipulations are necessary in order to arrive at a convincing interpretation of this effect. However, the fact that this effect occurred, suggests the value of looking more closely at subjects' strategies when dealing with non-Bernoulli sequences of stimuli.

2.4 Results of Experiment 2

2.4.1 Behavioral Results

Figures 2-5 and 2-6 present the behavioral results from Experiment 2. Figure 2-5 indicates that the accuracy of responding to targets, whether by overtly responding or by inhibiting an overt response, decreased with increasing workload. This trend was confirmed statistically. Figure 2-6 indicates that in both the "go on target" and "no go on target" tasks, reaction time increased with increasing number of readouts being attended. This trend, which also was highly statistically significant, replicates the findings from the behavioral test session of Experiment 1. Again, we can interpret these behavioral findings as indicating that the present workload manipulation was effective.

2.4.2 General Aspects of Obtained ERPs

The waveforms obtained in Experiment 2 differed somewhat between the "go on target" and "no-go on target" tasks, but in most important respects they replicated the effects seen in Experiment 1. Some of these differences were, no doubt, related to the fact that a motor potential was present in the waveforms elicited by stimuli that required an immediate behavioral response. For the present purposes, we focused primarily on determining the extent to

"GO ON TARGET" TASK

TRIALS ON WHICH TARGET OCCURED

TRIALS ON WHICH NON-TARGET OCCURED

"NO - GO ON TARGET" TASK

TRIALS ON WHICH TARGET OCCURED

TRIALS ON WHICH NON-TARGET OCCURED
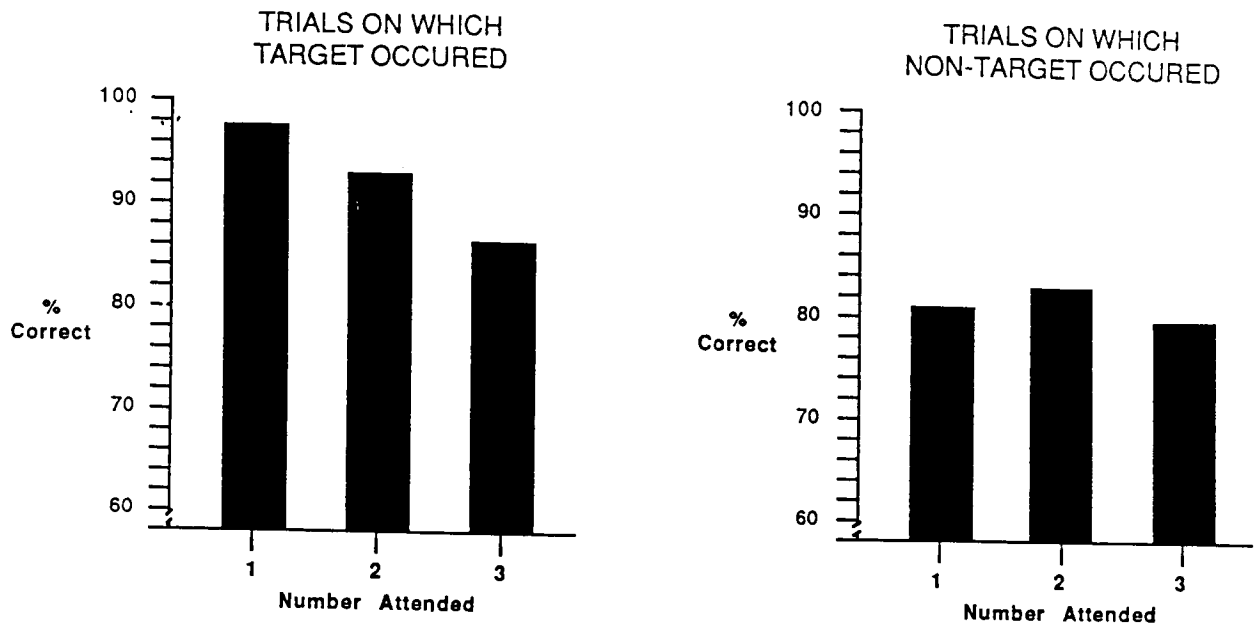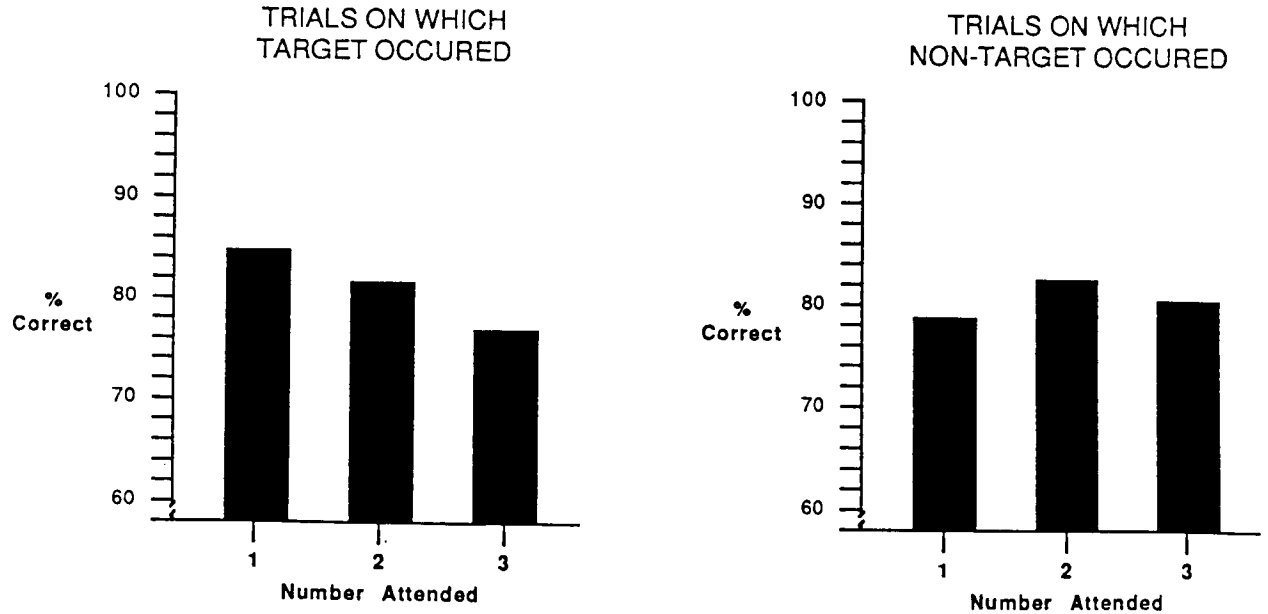
Figure 2–5.   Accuracy of responding, averaged across subjects, for the two response conditions in Experiment 2.

"GO ON TARGET" TASK
(Correct Responses Only)
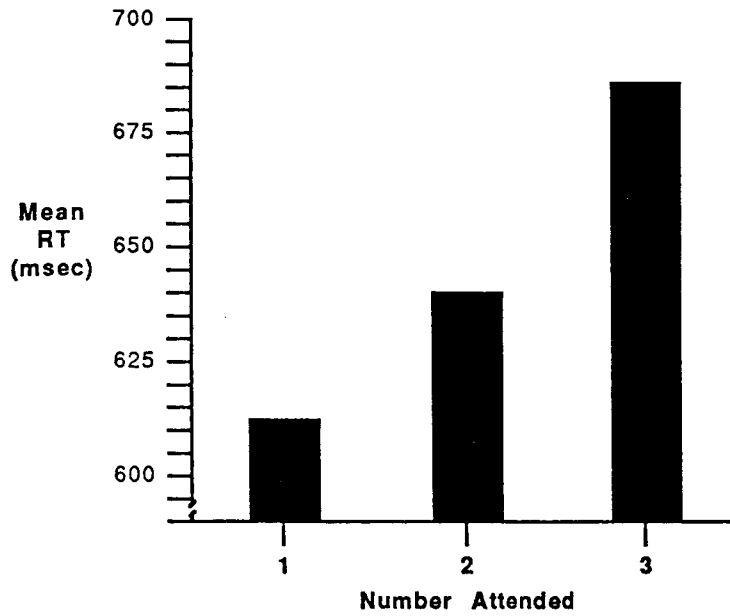
"NO - GO ON TARGET" TASK
(Correct Responses to Non-Targets Only)
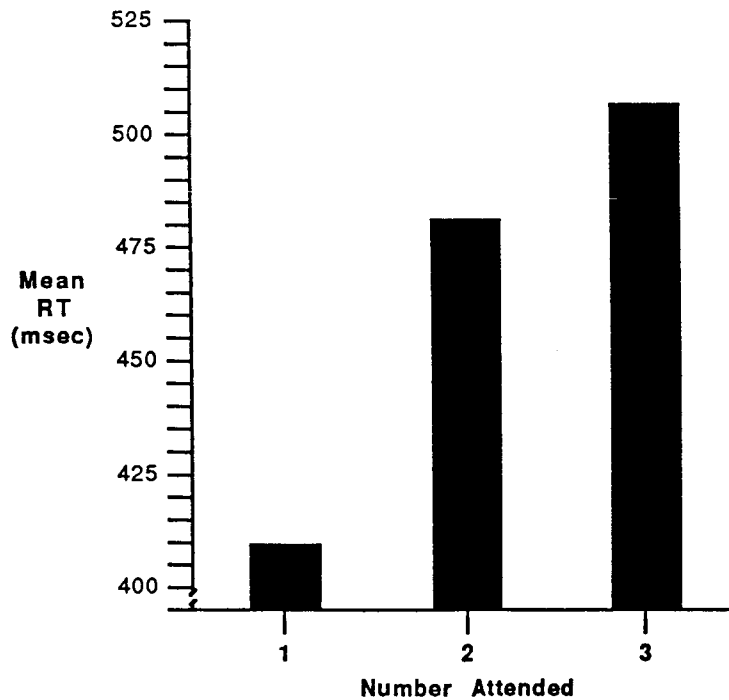
Figure 2-6. Mean reaction times, averaged across subjects, for the two response conditions in Experiment 2.

which the effects seen previously were replicated in the Experiment 2 waveforms that were elicited in the absence of a behavioral response (i.e., by "in bounds" readout changes that occurred during the "go on target" task and by "out-of-bounds" changes that occurred during the "no-go on target" task).

There were no effects of trend (directional bias) on either the behavioral measures or ERPs. Therefore, the ERP data were collapsed across trend. Figure 2-7 presents the raw, grand averaged waveforms from Experiment 2. Figure 2-8 presents the corresponding difference waveforms constructed by subtracting responses elicited under low workload conditions from those elicited under high workload conditions.

2.4.3 Target Effects

Again, a large increase in P300 amplitude was apparent in the responses elicited by target stimuli, i.e. changes in attended readouts that took the readout out-of-bounds. Interestingly, this target effect was clear no matter whether subjects responded to the target readouts with a motor response or the inhibition of a motor response. Also striking, particularly in the "no-go on target" data was the appearance of the workload-related N450 in the responses elicited by changes in target readouts. Both of these effects reached statistical significance in the data from both tasks.

2.4.4 Selective Attention Effects

The same attention-related effects that were seen in Experiment 1 were again found here. There were the same, difficult to interpret, differences in P300 between the responses elicited by attended and unattended readouts.

2.4.5 Tonic Workload Effects

The slow wave increases related to the number of readouts being monitored were again apparent under some conditions, but were somewhat less robust than in Experiment 1. Again the N250 wave was found consistently in the difference waveforms from all conditions when responses elicited under low workload were subtracted from the responses elicited under high workload.

COVER PAGE FOR FIGURES 2-7-A THROUGH 2-7-D
EXPERIMENT 2 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING TONIC WORKLOAD

Across-subject average waveforms from Experiment 2 at all mid-line scalp sites
and EOG.   Responses elicited under low workload (one readout being attended)
are overlaid with medium workload (two readouts being attended) and high
workload (three readouts being attended).   The left column contains trials on
which the subject made a finger movement to indicate his response, and the
right column contains trials on which the subject's lack of finger movement
indicated his response.   Responses are sorted according to whether the
eliciting change occurred in a monitored or non-monitored readout and whether
or not the eliciting change took the readout out-of-bounds.

2-7-A   Change in an unattended readout which remained in-bounds

2-7-B   Change in an attended readout which remained in-bounds

2-7-C   Change in an unattended readout which took the readout out-of-bounds

2-7-D   Change in an attended readout which took the readout out-of-bounds

EXPERIMENT 2 — RAW AVERAGE WAVEFORMS, ILLUSTRATING TONIC WORKLOAD
Figure 2-7-A. Change in an unattended readout which remained in-bounds

GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

NO-GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

S1

S1

15µV

650ms

ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT
WHICH KEPT THAT READOUT IN-BOUNDS, WHEN:

----------------3 HIGH WORKLOAD (THREE READOUTS BEING MONITORED)

·····················2 MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED)

———————1 LOW WORKLOAD (ONE READOUT BEING MONITORED)
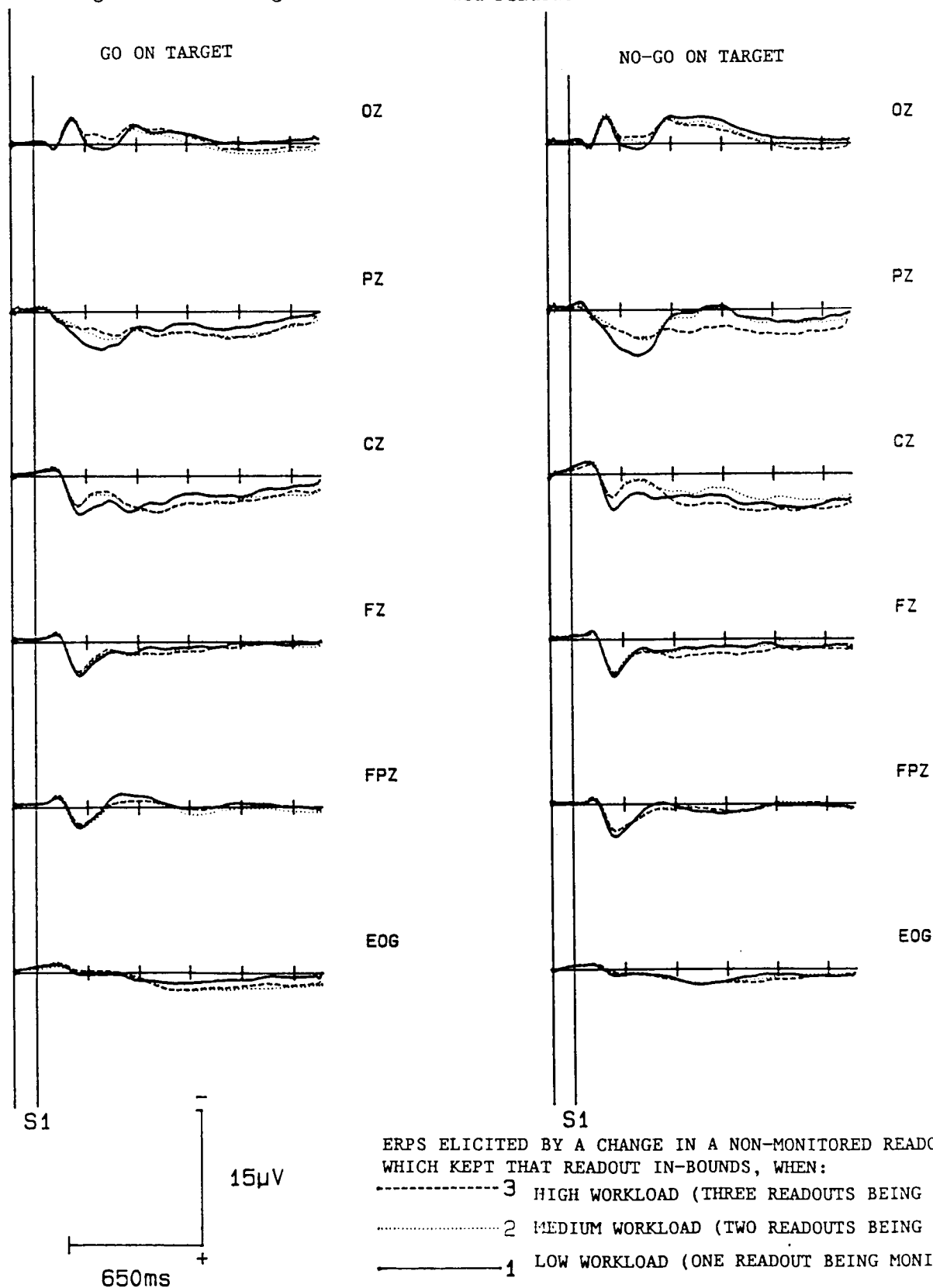
EXPERIMENT 2 — RAW AVERAGE WAVEFORMS, ILLUSTRATING TONIC WORKLOAD
Figure 2-7-B. Change in an attended readout which remained in-bounds

GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG   GO

S1

15μV

650ms

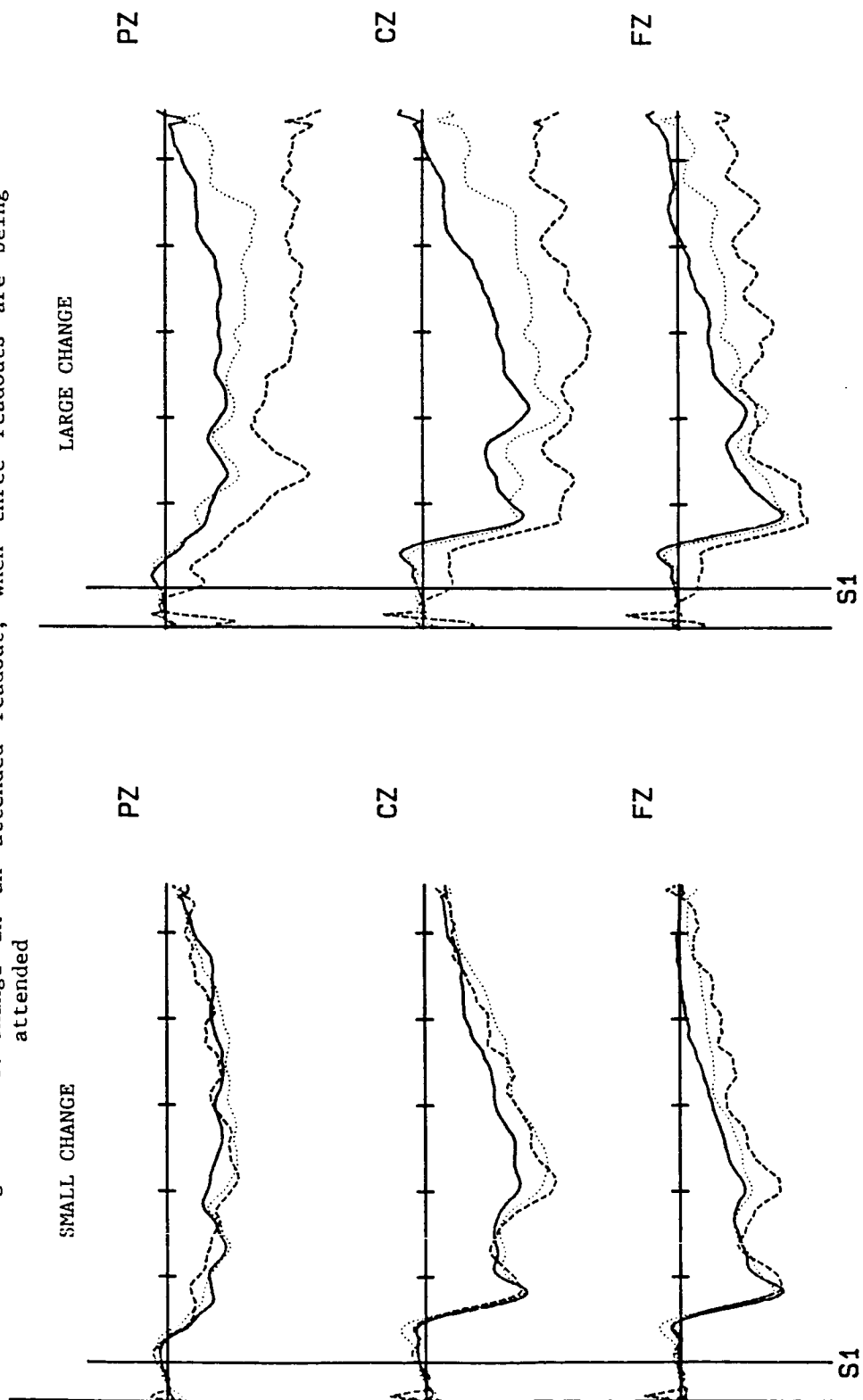NO-GO ON TARGET
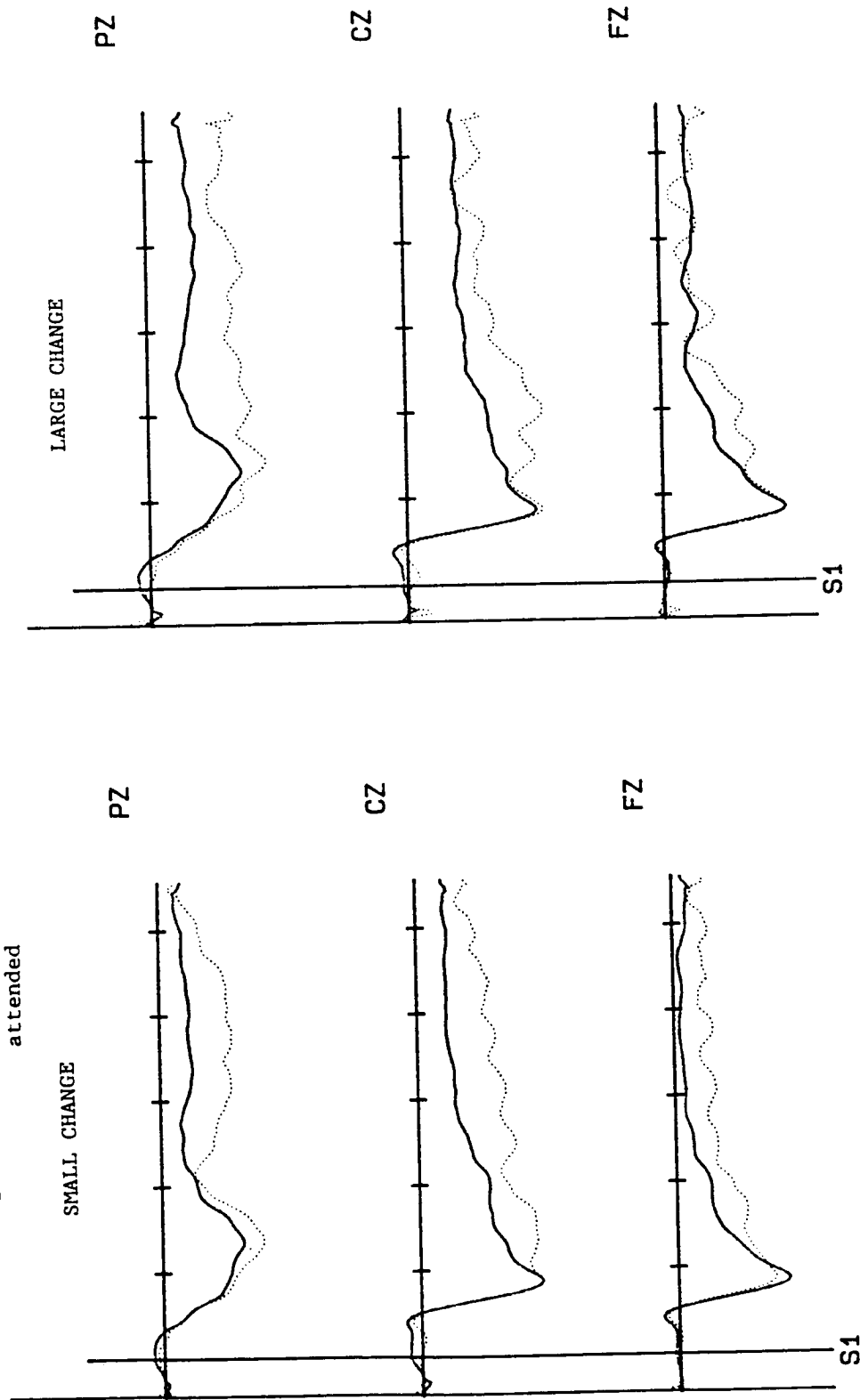
OZ

PZ

CZ

FZ

FPZ

EOG

S1

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT
WHICH KEPT THAT READOUT IN-BOUNDS, WHEN:

------------3 HIGH WORKLOAD (THREE READOUTS BEING MONITORED)

...................2 MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED)
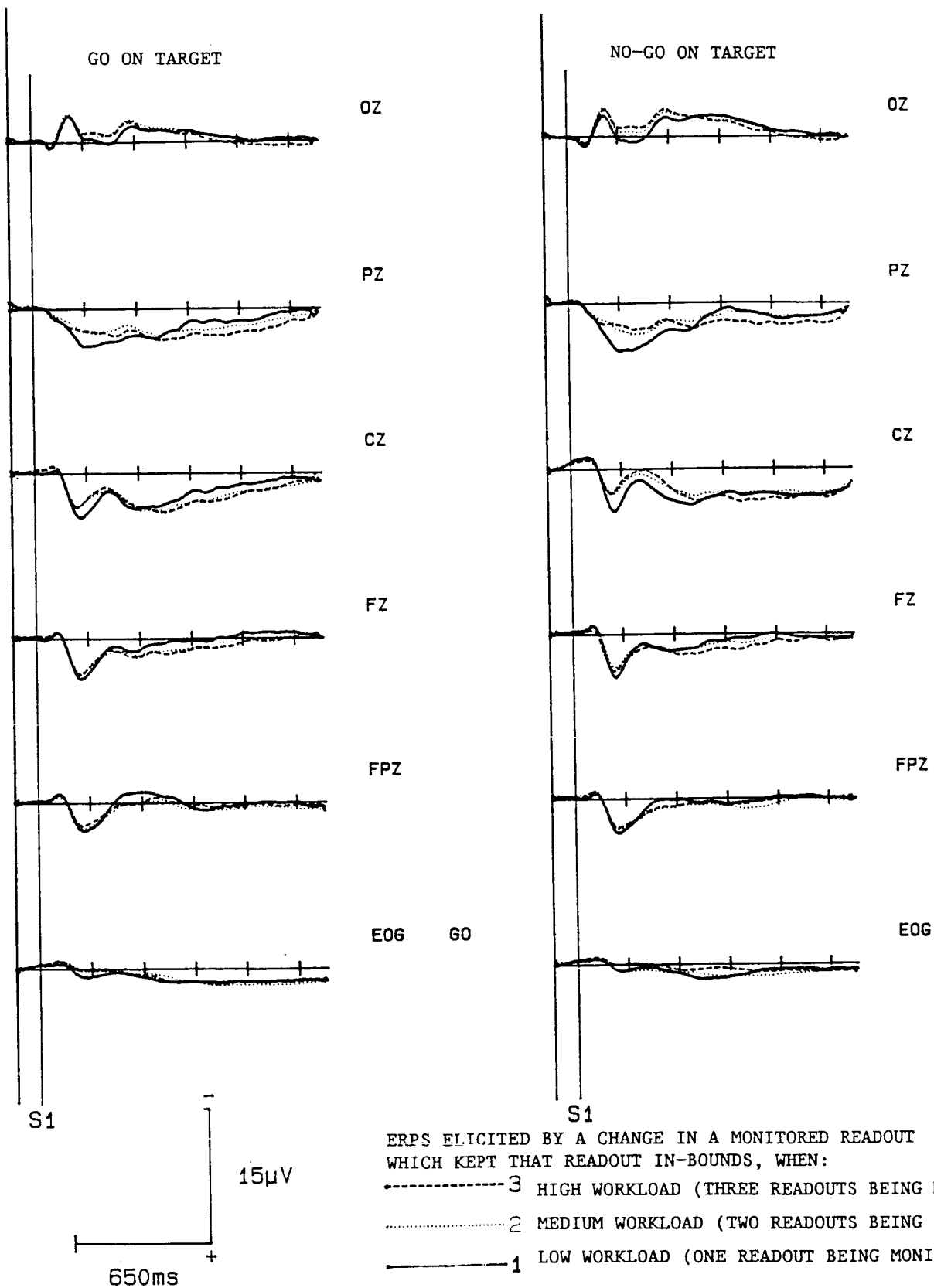
_____1 LOW WORKLOAD (ONE READOUT BEING MONITORED)

Figure 2-7-C. Change in an unattended readout which took the readout
out-of-bounds

GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

NO-GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

S1

S1

−

15μV

650ms

+

ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT
WHICH TOOK THAT READOUT OUT-OF-BOUNDS, WHEN:

--------------3 HIGH WORKLOAD (THREE READOUTS BEING MONITORED)

·····················2 MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED)

————————1 LOW WORKLOAD (ONE READOUT BEING MONITORED)

2-30
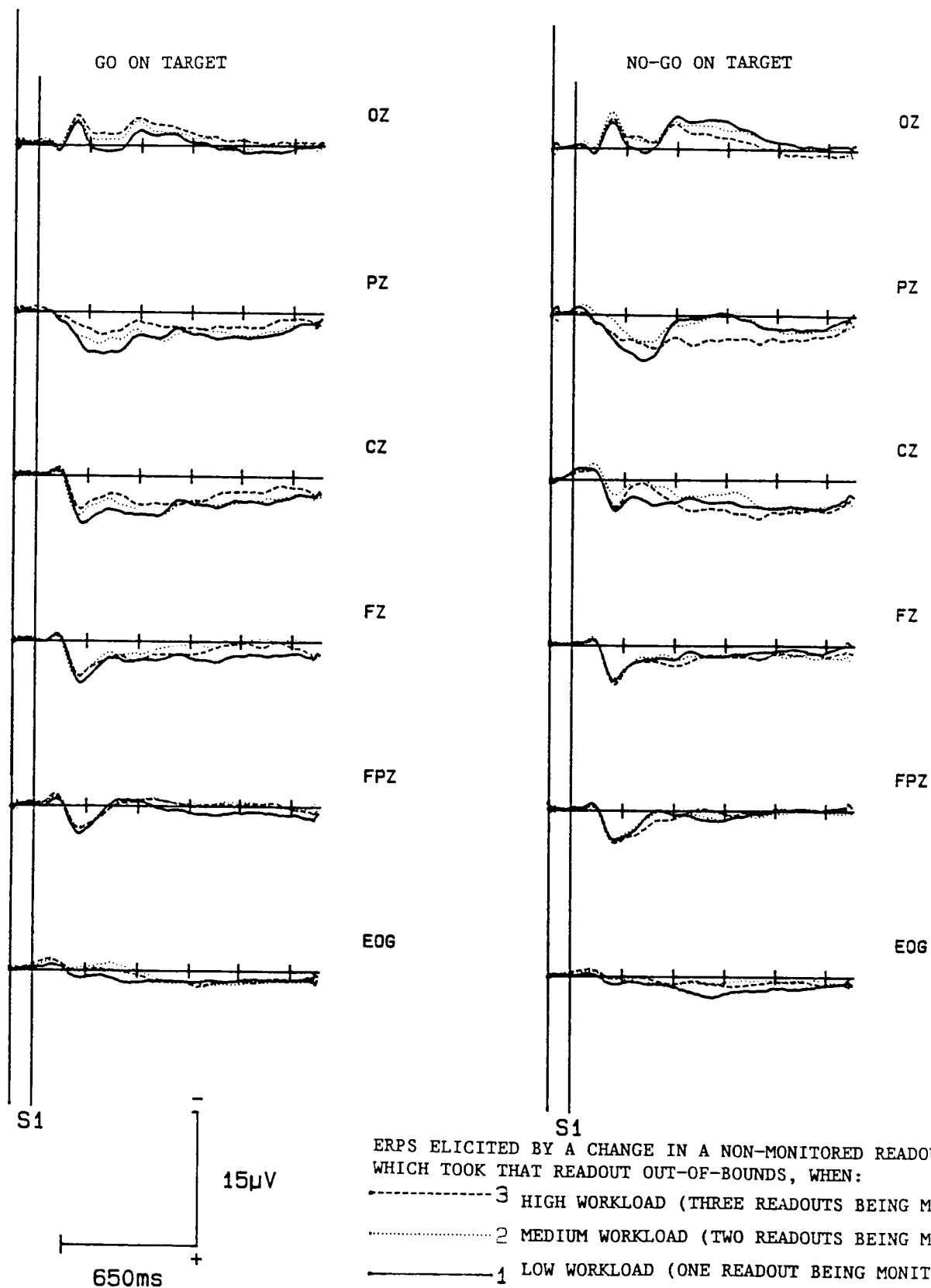
Figure 2-7-D. Change in an attended readout which took the readout out-of-bounds

GO ON TARGET

NO-GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

S1

15µV

650ms

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT
WHICH TOOK THAT READOUT OUT-OF-BOUNDS, WHEN:

--------------3 HIGH WORKLOAD (THREE READOUTS BEING MONITORED)

....................2 MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED)

_____1 LOW WORKLOAD (ONE READOUT BEING MONITORED)
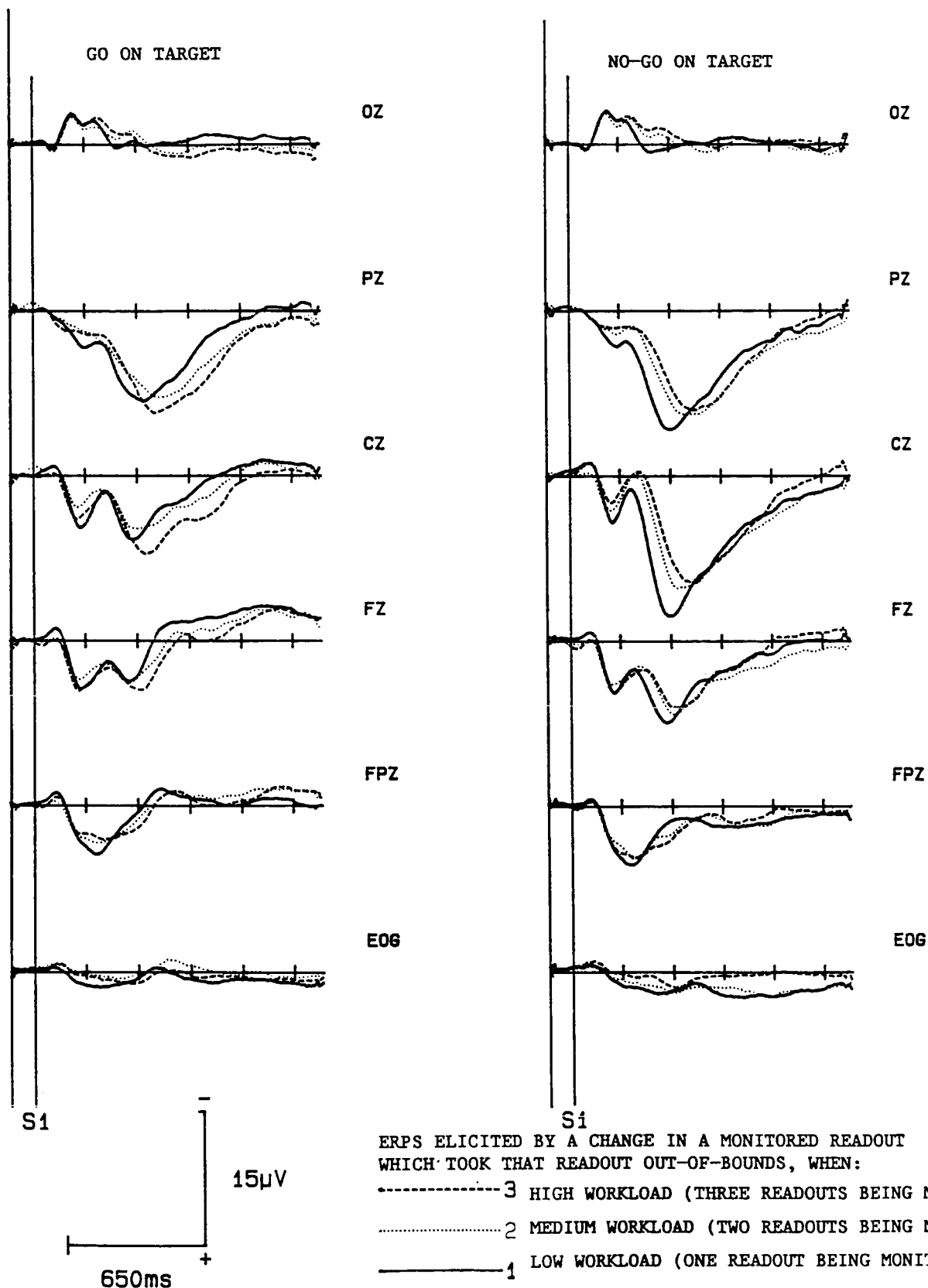
COVER PAGE FOR FIGURES 2-8-A THROUGH 2-8-D
EXPERIMENT 2 -- DIFFERENCE WAVEFORMS, ILLUSTRATING TONIC WORKLOAD

Difference waveforms corresponding to the data in Figure 2-7, with the responses elicited under low workload conditions subtracted from the responses elicited under high workload.

2-8-A   Change in an unattended readout which remained in-bounds

2-8-B   Change in an attended readout which remained in-bounds

2-8-C   Change in an unattended readout which took the readout out-of-bounds

2-8-D   Change in an attended readout which took the readout out-of-bounds

EXPERIMENT 2 — DIFFERENCE WAVEFORMS, ILLUSTRATING TONIC WORKLOAD
Figure 2-8-A. Change in an unattended readout which remained in-bounds

GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

S1

NO-GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

S1

15μV

650ms

ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT
WHICH KEPT THAT READOUT IN-BOUNDS
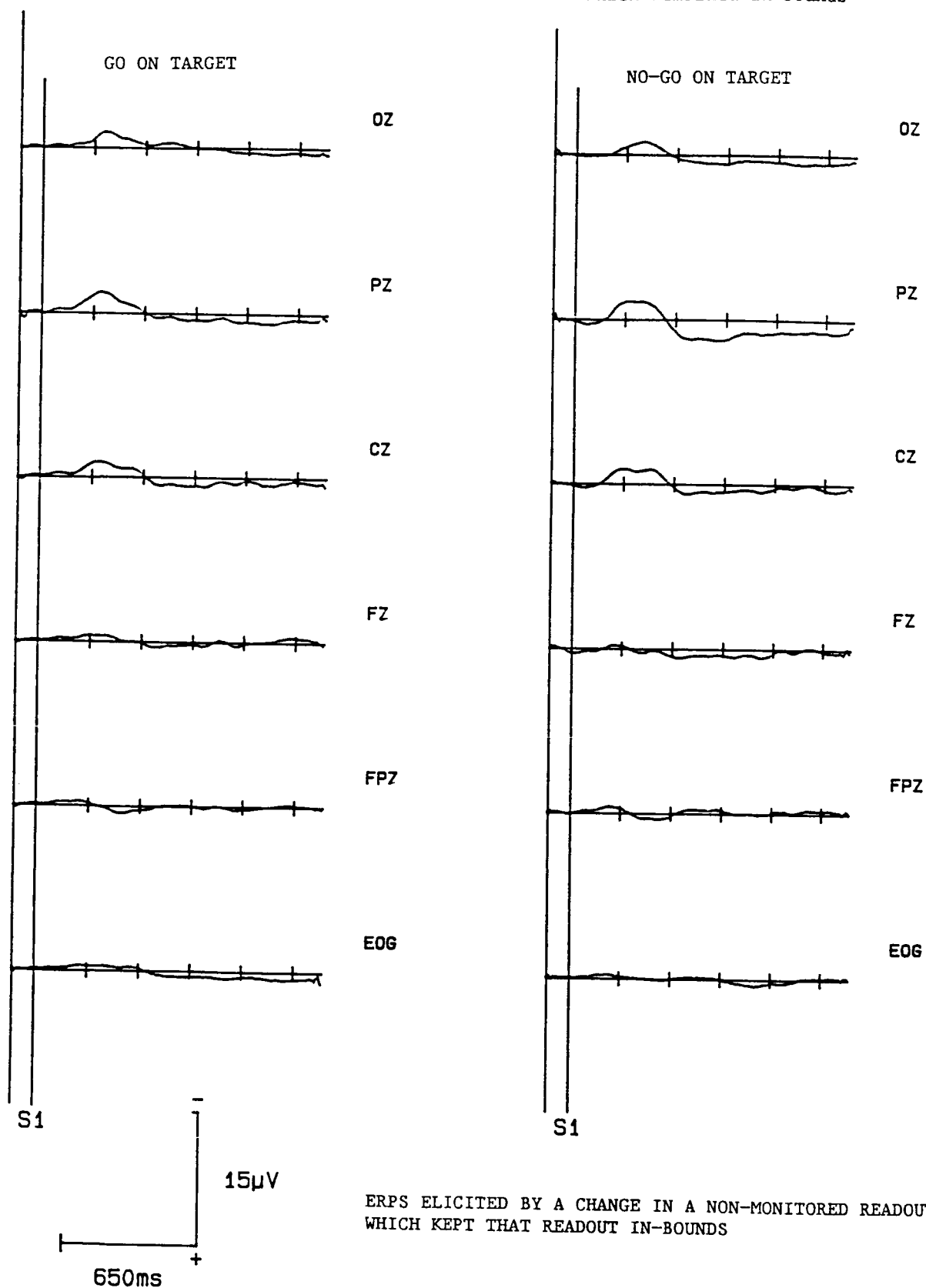
EXPERIMENT 2 — DIFFERENCE WAVEFORMS, ILLUSTRATING TONIC WORKLOAD
Figure 2-8-B. Change in an attended readout which remained in-bounds

GO ON TARGET

OZ

PZ

CZ

FZ

FP7

EOG

S1

15μV

650ms

NO-GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

S1

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT
WHICH KEPT THAT READOUT IN-BOUNDS
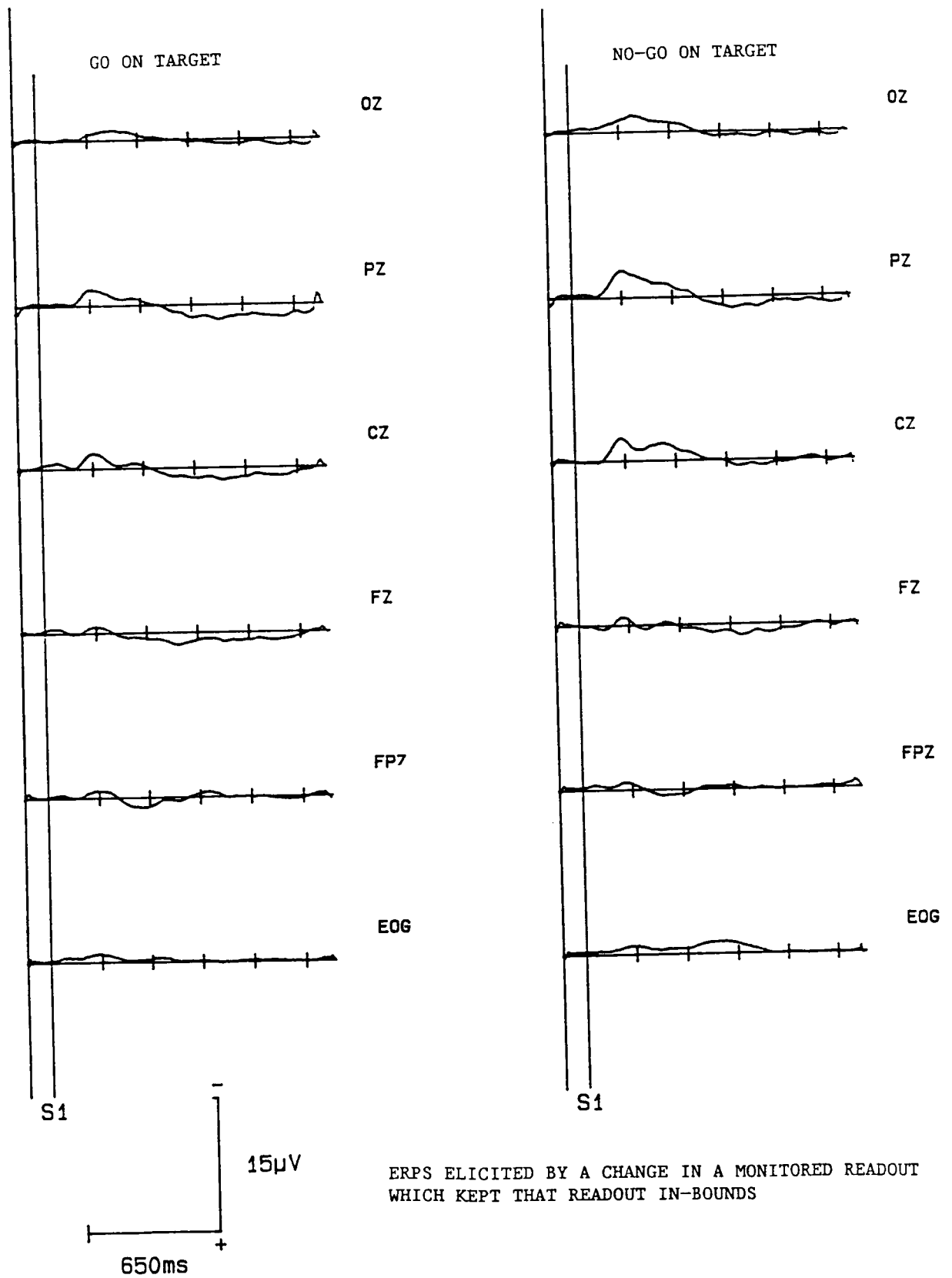
EXPERIMENT 2 — DIFFERENCE WAVEFORMS, ILLUSTRATING TONIC WORKLOAD
Figure 2-8-C. Change in an unattended readout which took the readout out-of-bounds
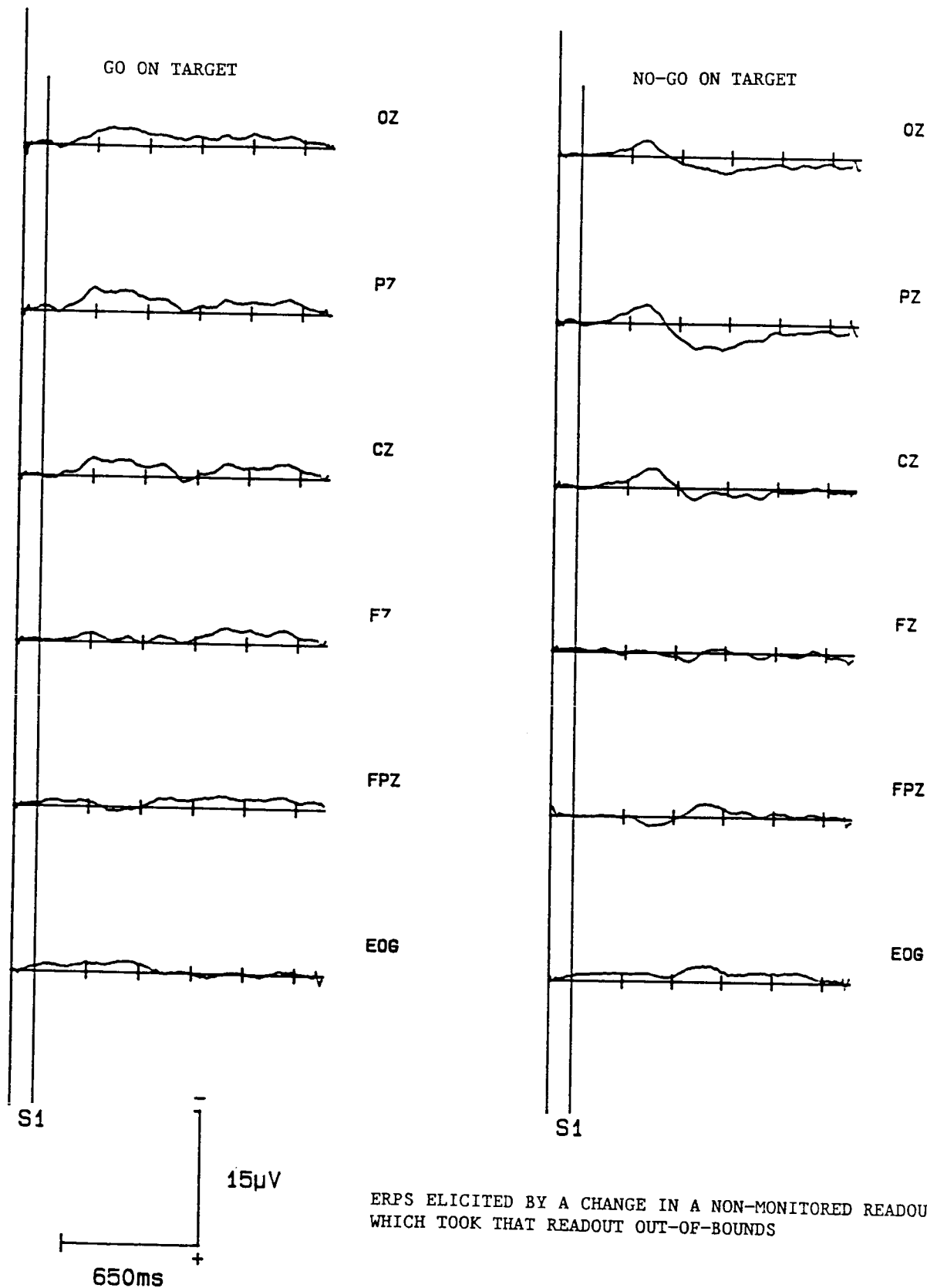
GO ON TARGET

OZ
P7
CZ
F7
FPZ
EOG

S1

NO-GO ON TARGET

OZ
PZ
CZ
FZ
FPZ
EOG

S1

15µV

650ms

ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT
WHICH TOOK THAT READOUT OUT-OF-BOUNDS

2-35

EXPERIMENT 2 — DIFFERENCE WAVEFORMS, ILLUSTRATING TONIC WORKLOAD

Figure 2-8-D. Change in an attended readout which took the readout out-of-bounds

GO ON TARGET

NO-GO ON TARGET

OZ

PZ

CZ

FZ

FPZ

EOG

S1

15μV

650ms

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT
WHICH TOOK THAT READOUT OUT-OF-BOUNDS

2-36
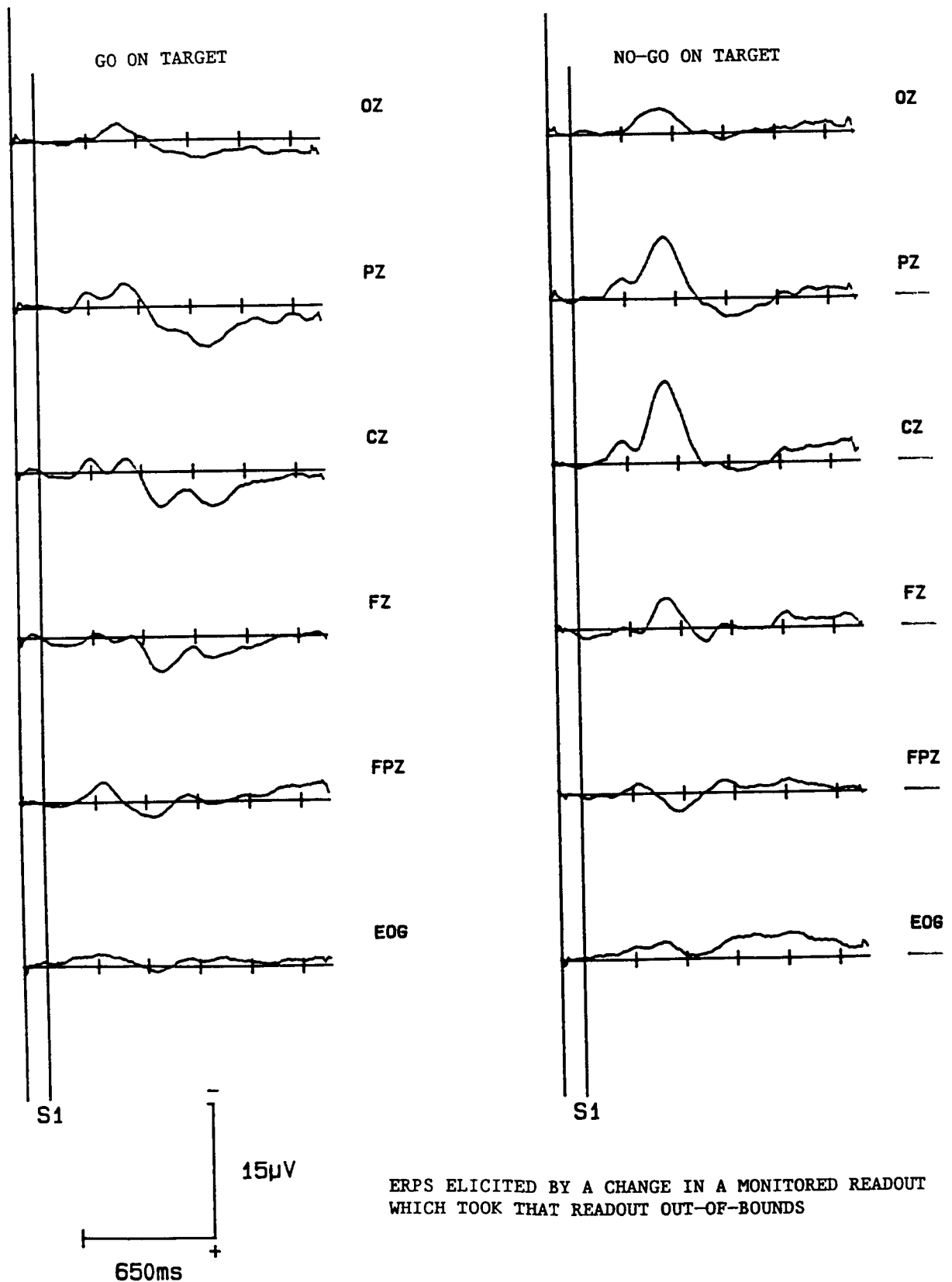
2.4.6 Phasic Effects of the Number of Readouts in Danger

Figure 2-9 presents the waveforms obtained when the data from the various conditions of Experiment 2 were sorted according to the number of attended readouts "in danger." Again, there were increases in the amplitude of the late positivity when increasing numbers of attended readouts were close to going out-of-bounds.

2.5 Discussion

Obviously, the monitoring task that we designed provided a rich environment for eliciting cognition-related effects on scalp-recorded ERPs. To summarize, the ERPs collected here were characterized by the following features:

o   An N250 wave, possibly a Processing Negativity (e.g., Naataanen, 1982), that emerged with increasing workload, in the responses to all readouts.

o   An N450 wave, possibly related to the N2 complex (e.g., Ritter, Simson, & Vaughan, 1983), that emerged with increasing workload, in responses to the target stimuli only.

o   A peak positivity, probably related to the P300 (e.g., Donchin, et al., 1986), which dramatically increased in amplitude when a target stimulus occurred, increased in amplitude as a function of the number of monitored readouts "in danger," and showed an interaction with tonic workload and selective attention, such that the differences between responses to monitored and non-monitored readouts which were found at low workload levels diminished with the requirement to monitor more readouts.

o   A slow positivity, possibly related to the Slow Wave (e.g., Ruchkin & Sutton, 1983), which increased in amplitude with workload, in the responses to all readouts.

More work is required to determine the functional significance of the waveform changes we observed and to relate them convincingly to ERP components that have been identified in other paradigms. Nevertheless, the present findings warrant

COVER PAGE FOR FIGURES 2-9-A THROUGH 2-9-H
EXPERIMENT 2 — RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Across-subject average waveforms from Experiment 1 at Pz, Cz, and Fz, with responses elicited when different numbers of attended readouts were "in danger," i.e., within an incremental value of going "out-of-bounds." The left column contains trials on which the change did not take the readout out-of-bounds, and the right column contains trials on which the change did take the readout out-of-bounds. The responses are sorted according to whether the eliciting change occurred in a attended or an unattended readout, the number of readouts being attended at the time, and whether the subjects finger movement indicated his response (go), or whether the subjects lack of finger movement indicated his response (no/go).

2-9-A   Change in an attended readout, when two readouts are being attended, subject's finger movement indicates his response

2-9-B   Change in an attended readout, when three readouts are being attended, subject's finger movement indicates his response

2-9-C   Change in an unattended readout, when two readouts are being attended, subject's finger movement indicates his response

2-9-D   Change in an unattended readout, when three readouts are being attended, subject's finger movement indicates his response

2-9-E   Change in an attended readout, when two readouts are being attended, subject's lack of finger movement indicates his response

2-9-F   Change in an attended readout, when three readouts are being attended, subject's lack of finger movement indicates his response

2-9-G   Change in an unattended readout, when two readouts are being attended, subject's lack of finger movement indicates his response

2-9-H   Change in an unattended readout, when three readouts are being attended, subject's lack of finger movement indicates his response

EXPERIMENT 2 — RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-9-A. Change in an attended readout, when two readouts are being attended, subject's finger movement indicates his response

GO ON TARGET TASK

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT,
WHEN MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED),
AND WHEN:

- - - - - - - - - - 2 MONITORED READOUTS IN DANGER
.... 1 MONITORED READOUT IN DANGER
————————— 0 MONITORED READOUTS IN DANGER
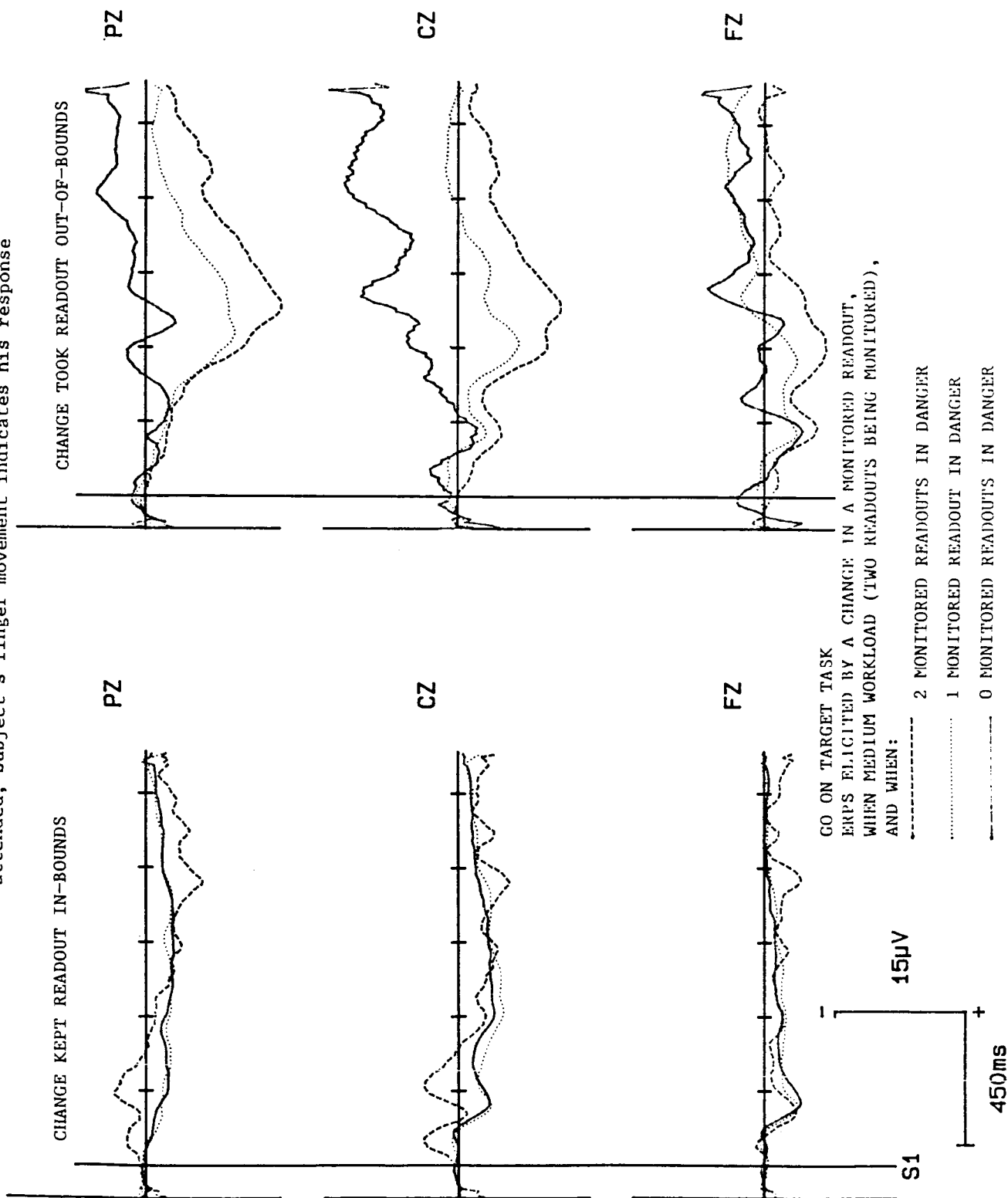
EXPERIMENT 2 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-9-B. Change in an attended readout, when three readouts are being attended, subject's finger movement indicates his response

GO ON TARGET TASK

ERPS ELICITED BY A CHANGE
IN A MONITORED READOUT, WHEN HIGH WORKLOAD (THREE READOUTS BEING MONITORED),
AND WHEN:

- - - - - - - 2 MONITORED READOUTS IN DANGER

.... 1 MONITORED READOUT IN DANGER

- - - - - - 0 MONITORED READOUTS IN DANGER
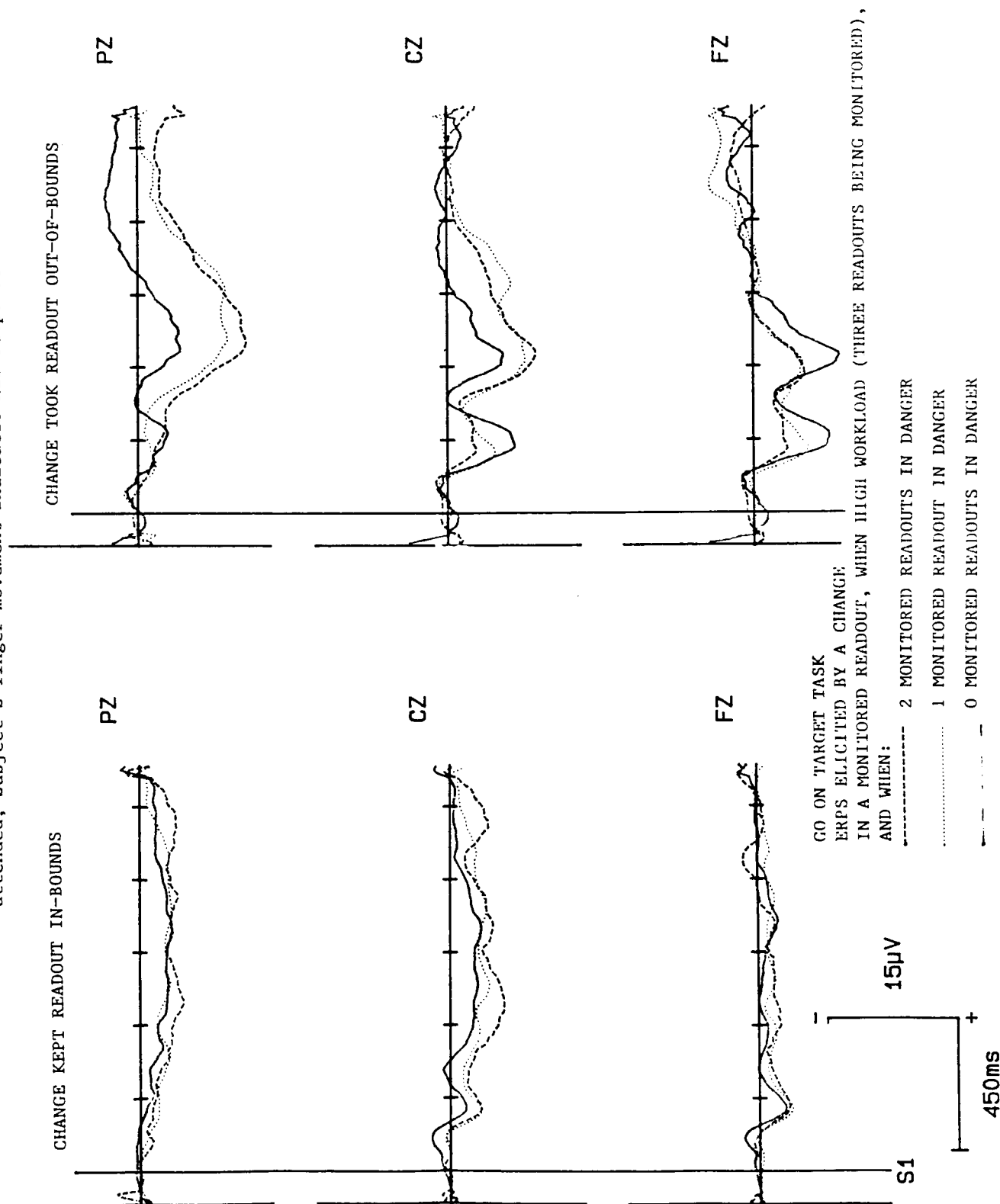
EXPERIMENT 2 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-9-C. Change in an unattended readout, when two readouts are being attended, subject's finger movement indicates his response

CHANGE TOOK READOUT OUT-OF-BOUNDS

PZ

CZ

FZ

CHANGE KEPT READOUT IN-BOUNDS

PZ

CZ

FZ

GO ON TARGET TASK
ERP'S ELICITED BY A CHANGE IN A NON-MONITORED READOUT,
WHEN MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED),
AND WHEN:

——————— 2 MONITORED READOUTS IN DANGER

··············· 1 MONITORED READOUT IN DANGER

—·—·—· 0 MONITORED READOUTS IN DANGER

15μV

+

450ms

S1
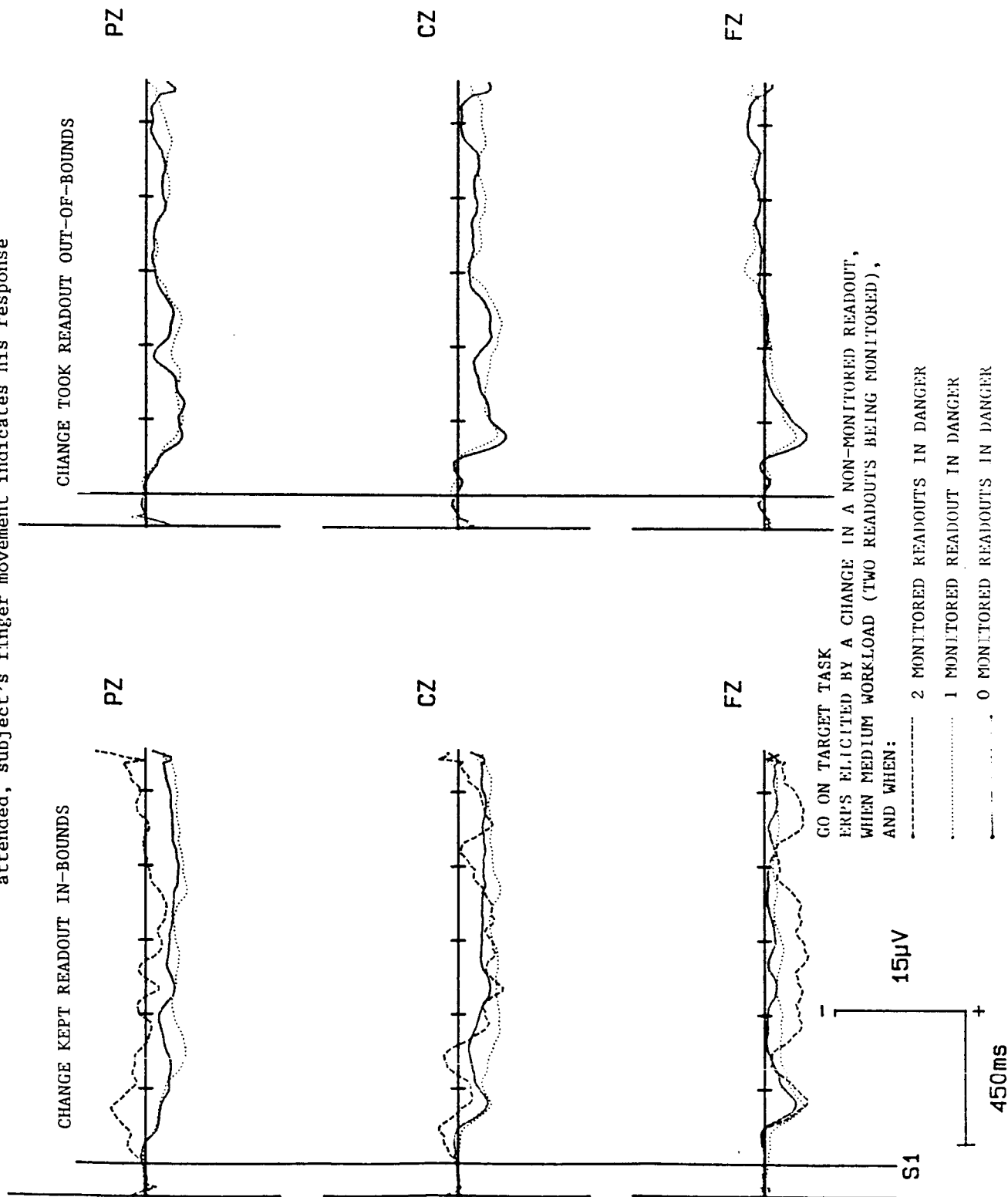
EXPERIMENT 2 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS
Figure 2-9-D. Change in an unattended readout, when three readouts are being attended, subject's finger movement indicates his response

CHANGE TOOK READOUT OUT-OF-BOUNDS

CHANGE KEPT READOUT IN-BOUNDS

PZ

CZ

FZ

GO ON TARGET TASK
ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT,
WHEN HIGH WORKLOAD (THREE READOUTS BEING MONITORED),
AND WHEN:

--------- 2 MONITORED READOUTS IN DANGER
.... 1 MONITORED READOUT IN DANGER
___ 0 MONITORED READOUTS IN DANGER

15µV

S1

EXPERIMENT 2 -- RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-9-E. Change in an attended readout, when two readouts are being attended, subject's lack of finger movement indicates his response

CHANGE KEPT READOUT IN-BOUNDS

CHANGE TOOK READOUT OUT-OF-BOUNDS

PZ

PZ

CZ

CZ

FZ

FZ

15μV

450ms

S1

NO-GO ON TARGET TASK

ERPS ELICITED BY A CHANGE IN A MONITORED READOUT, WHEN MEDIUM WORKLOAD AND WHEN:

—————— 2 MONITORED READOUTS IN DANGER

————— 1 MONITORED READOUT IN DANGER

·········· 0 MONITORED READOUTS IN DANGER

2-43
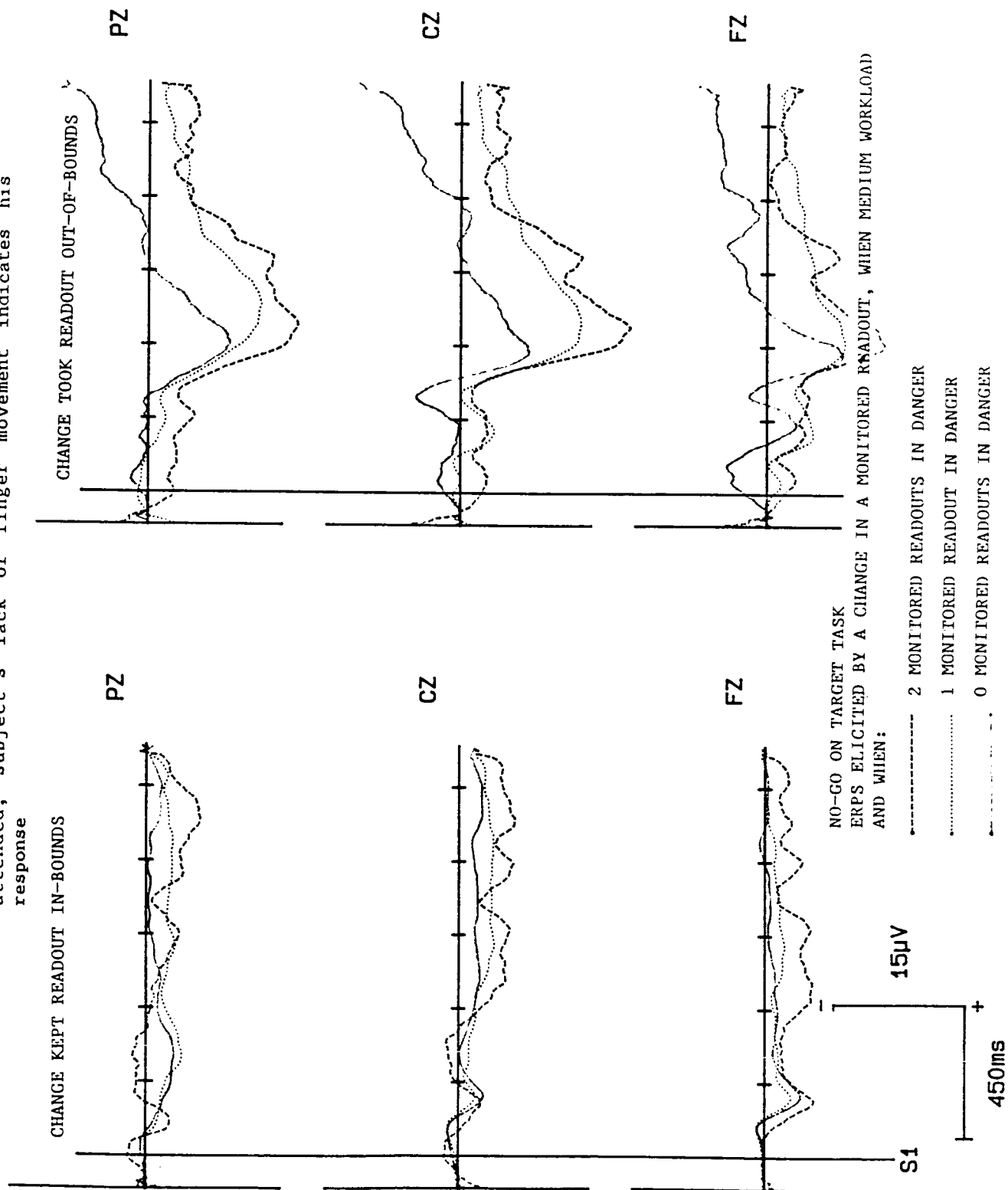
EXPERIMENT 2 — RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-9-F. Change in an attended readout, when three readouts are being attended, subject's lack of finger movement indicates his response

CHANGE TOOK READOUT OUT-OF-BOUNDS

PZ

CZ

FZ

CHANGE KEPT READOUT IN-BOUNDS

PZ

CZ

FZ

NO-GO ON TARGET TASK
ERPS ELICITED BY A CHANGE IN A MONITORED READOUT,
WHEN HIGH WORKLOAD (THREE READOUTS BEING MONITORED),
AND WHEN:

— 2 MONITORED READOUTS IN DANGER
.... 1 MONITORED READOUT IN DANGER
— 0 MONITORED READOUTS IN DANGER

15μV

+

−

450ms

S1

EXPERIMENT 2 — RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-9-G. Change in an unattended readout, when two readouts are being attended, subject's lack of finger movement indicates his response

PZ

CZ

FZ

CHANGE TOOK READOUT OUT-OF-BOUNDS

PZ

CZ

FZ

CHANGE KEPT READOUT IN-BOUNDS

15μV

450ms

S1

NO-GO ON TARGET TASK

ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT, WHEN MEDIUM WORKLOAD (TWO READOUTS BEING MONITORED), AND WHEN:

— 2 MONITORED READOUTS IN DANGER

···· 1 MONITORED READOUT IN DANGER

— — 0 MONITORED READOUTS IN DANGER

EXPERIMENT 2 — RAW AVERAGE WAVEFORMS, ILLUSTRATING PHASIC EFFECTS

Figure 2-9-H. Change in an unattended readout, when three readouts are being attended, subject's lack of finger movement indicates his response

CHANGE KEPT READOUT IN-BOUNDS

CHANGE TOOK READOUT OUT-OF-BOUNDS

PZ IN   PZ OUT

CZ   CZ OUT

FZ   FZ OUT

15µV

450ms

S1

NO-GO ON TARGET TASK

ERPS ELICITED BY A CHANGE IN A NON-MONITORED READOUT, WHEN HIGH WORKLOAD (THREE READOUTS BEING MONITORED), AND WHEN:

—————— 2 MONITORED READOUTS IN DANGER

············· 1 MONITORED READOUT IN DANGER

— · — · — 0 MONITORED READOUTS IN DANGER

several important general conclusions:

o   Workload-related ERP effects can be derived in single-task paradigms without burdening the subject with competing task demands.

o   The effects of different cognitive variables are specific to circumscribed regions of the waveforms.

o   Some regions of the waveforms are affected by multiple information-processing manipulations.

These relationships confirm the exquisite sensitivity of scalp-recorded ERPs to the cognitive milieu in demanding tasks and suggest the possibility of eventually indexing specific cognitive processes with specific waveform components or with the activity in specific latency regions of ERPs.

It is interesting to note, however, that even prior to attaining a thorough understanding of the functional significance of specific ERP components, one can infer, from the pattern of results, a number of indications about how subjects performed the present task.   Consider the fact that changes in monitored readouts that went out-of-bounds (i.e. targets) elicited a markedly different response from changes in monitored readouts that stayed in-bounds, whereas responses to changes in non-monitored readouts did not distinguish between in-bounds and out-of-bounds changes.   These results suggest that subjects did indeed selectively attend to the readout positions that they were instructed to monitor.   Likewise, the fact that the ERPs showed a significant effect related to the number of monitored readouts "in danger," but no effect of the number of non-monitored readouts "in danger," suggests that subjects noticed the former but not the latter.   Both of these findings are consistent with the conclusion that subjects did not process the value of non-monitored readouts, despite the fact that only one readout changed on a given presentation and subjects did not know whether a monitored or non-monitored readout was about to change.

On the other hand, this conclusion must be reconciled with the fact that both the workload effect on the N250 and slow positivity, and the effect of number

of monitored readouts "in danger" on the peak positivity, were found in the responses to changes in both monitored and non-monitored readouts. This finding suggests that these ERP effects reflect differential processing due to the distributing of attention among the readouts being monitored, and that this processing, in essence, is related to determining which readout changed, rather than to determining the specific value of the readout that changed. Therefore, the present ERP results can be used to infer that subjects selectively attended to the readouts that they were to monitor, that they noticed the number of monitored readouts that were "in danger" of going out-of-bounds, and that workload modified some aspects of the processing of all stimuli, whether monitored or not.

Such information would be useful to know in a number of practical applications. Design issues such as configuring display formats which minimize workload, maximizing the effectiveness of warning messages, and increasing the salience of task-critical information often hinge on reliable measures of which stimuli are being attended, whether extraneous information is intrusive, whether subjects are taking advantage of useful information that is available, and which of several alternative designs entail less mental workload. The present results point towards the possibility of using ERPs to address such issues, in situations where one can not rely on, or it is difficult to acquire, subjective and behavioral measures. Moreover, in addition to playing a confirmatory or surrogate role, ERPs may serve a diagnostic function. When overt performance has been observed to fail, one may be able to glean information from ERP effects like those obtained here in order to indicate the particular aspects of information-processing, and by inference the particular aspects of system design, that were deficient. Beyond the design arena, such ERP measures may also be helpful for monitoring the progress of training on demanding tasks or for selecting personnel who are particularly capable of functioning in various tasks.

Of course, many of the ERP effects obtained here were small and required extensive data analysis based on average waveforms. For some engineering applications, one would have the luxury of collecting as much data and analyzing it to the extent that we did here, but in other applications one would be more constrained. Nonetheless, the present results may point the way

towards other manipulations or measures that would better emphasize the effects of interest. It will be interesting to see, as studies like the present ones are recast into operational systems or simulators whose task demands have been approximated in the laboratory, to what extent the cognitive-related patterns of ERP results become more pronounced.

3.0 TASK ANALYSIS OF ADVANCED CONCEPTS FLIGHT SIMULATOR DISPLAYS AND AIRCREW TASKS

Previous attempts to record physiological indices related to mental workload in-flight or in a cockpit simulator have usually introduced contrived stimuli and tasks such as those used in the laboratory. For example, in studies of scalp-recorded event-related potentials (ERPs), subjects might be given a "secondary" tone counting task to be performed while they control the aircraft or simulator. ERPs elicited by, for example, rare and frequency tones are then examined as some aspect of "primary" task workload, i.e. the difficulty of the flying task, is varied. This approach has the obvious advantages of providing discrete stimuli (the tones) to which the ERPs can be time-locked and allowing the triggering of these recordings with instrumentation that is independent of the cockpit instruments and controls. However, this approach clearly has very limited applicability in the context of realistic scenarios. The tone counting task is obtrusive in that it burdens pilots with additional workload as compared to that which they would otherwise experience. Therefore, conclusions drawn from such contrived situations are of questionable validity and limited generality.

3.1 Objectives of the Present Analysis

If ERPs are to find a place in the system design process, they will have to be recorded unobtrusively in response to events to which the pilot will be exposed in the course of normal flight operations. The objective in the present effort was to take a first step in this direction by conducting an analysis of cockpit displays and pilot tasks, for the purpose of identifying cockpit events which might be expected to elicit ERPs related to pilot mental workload.

The Advanced Concepts Flight Simulator (ACFS) at NASA Langley provided the context in which to conduct this analysis. This state-of-the-art simulator, designed to model a 1995 time-frame transport aircraft, was being implemented during the present period of performance as a research simulator at NASA Langley, NASA Ames and Lockheed-Georgia. It provided an attractive framework

in which to address the possibility of ERP recordings, because it makes extensive use of CRT displays to convey flight information to the aircrew, these displays are well documented, it will support a wide range of full-task mission scenarios in which workload can be manipulated, and because the aircraft it foreshadows will probably be available during the time-frame within which one might hope ERP technology will find its way into operational systems. Moreover, a number of methodological issues that must be addressed in considering the recording of ERPs in the ACFS are typical of those that will be faced by investigators working with other simulators.

The present effort focused on the following issues:

o   How to identify events in the cockpit that should elicit ERPs related to workload. The dual goals were to develop a methodology that could also be applied for such purposes in other systems and to specifically identify ACFS events that could be used in a subsequent validation effort to test hypotheses about ERP indices of workload under realistic conditions.

o   What ACFS tasks and scenarios would likely impose high workload demands on the aircrew and how to manipulate workload during realistic scenarios in the ACFS. In that it models an aircraft that is presently somewhat beyond the state-of-the-art, the ACFS incorporates a number of automated capabilities that were designed to reduce the information processing load on the aircrew. Nonetheless, the continuing expansion of instrumentation and sensors being forecast for near-term implementation in cockpits ensure that, at least under certain operational conditions, high mental workload will continue to be an operational problem.

o   How to derive converging measures of workload that can be correlated with ERP measures, again without resorting to contrived or overly obtrusive indices. Such converging measures are important, both in order to confirm that workload is indeed being manipulated in experimental efforts and to explore the extent to which ERP measures correlate with other indices of workload.

These issues were addressed using modified task analysis techniques. In most

task analyses, the human factors engineer takes great care to characterize the operator's tasks in terms of observable stimuli and consequent actions. The present approach was, by necessity, somewhat different. We wished to characterize the pilot's tasks in terms of the psychological constructs that have been related to ERP activity -- selective attention, expectancy for specific events, uncertainty resolution, workload and spare capacity to process incoming stimuli, and stimulus discriminability (for example, see reviews by Donchin, Ritter, & McCallum, 1978; Donchin, Kramer, & Wickens, 1986). By analyzing the pilot's tasks and cockpit displays in terms of such inferred mental constructs, it was possible to predict the particular cockpit events, occurring in specific contexts, to which it may be fruitful to time-lock ERP recordings for the purpose of assessing pilot workload.

3.2 Analysis Methods

As mentioned previously, one reason for selecting the ACFS for the present analysis was the fact that there was extensive conceptual design documentation available (Sexton & Needles, 1982; Sexton, 1983). Drawing upon the expertise of subject matter experts at Lockheed-Georgia and NASA Langley, ARD human factors engineers first familiarized themselves with the key "baseline" displays planned for the ACFS (see Addendum 3A-1), as documented in Sexton, 1983. Each display was then analyzed in turn, with the tasks during which each would be utilized being documented. The display elements for each display were characterized according to the following criteria, which addressed the extent to which each element was a candidate for our purposes:

o   The extent to which the display element changes discretely. In order to elicit ERPs, using conventional recording techniques, one must be able to time-lock to a discrete point in time at which the event begins. This requirement need not imply the appearance of a stimulus (e.g. on a CRT) which had been absent prior to that time, but it does imply a discrete, non-continuous change in the display element. For the time being, we did not consider how to provide the appropriate timing information.

o   The extent to which the display element delivers information to the aircrew in a manner that might elicit ERP components related to workload.

Candidate display elements were examined in the context of loosely defined tasks to determine the extent to which they would elicit the mental processes that have been associated with the P300 and other late positive components of the ERP. In particular, does the display element convey information that is somewhat unexpected or uncertain, is it readily discriminable (e.g. with respect to clutter on the CRT or to its distinctiveness relative to other changes that could occur), and does it command attention, at least during certain tasks? It should be noted that the interest here was not in all display changes that are discrete enough to elicit an ERP or in all those that afford the opportunity for time-locking a recording. Rather, we were interested specifically in those ERP-eliciting display changes for which the ERP would be expected to contain the endogenous components that have been shown to vary with workload.

o   the extent to which the display or display element would likely be attended in situations that lend themselves to the manipulation of task difficulty and to ERP recordings. This criterion obviously required some hypothesizing about aircrew tasks and sources of workload in the ACFS. It also raised the question of whether the candidate changes in displays could realistically be presented repetitively, in order to perform conventional signal averaging of ERPs over at least some small number of occurrences.

A question that arose in the course of the above analysis was the likelihood that pilots would be attending to a given display when the candidate ERP-eliciting change occurred. This likelihood would clearly depend on such factors as the importance of the task which involved the display or display element of interest, the number and importance of competing tasks, and individual differences in pilots' style and preferences. In order to better determine likely pilot attentional patterns, and to work towards defining flight scenarios that would be useful for validating the speculated workload influence on candidate ERPs, a more detailed task analysis of one particular flight scenario was conducted.

An approach and landing scenario (see Addendum 3A-2) was adopted for this purpose, because it seemed to encompass most of the candidate display features

identified in the foregoing analysis. The scenario was broken down into detailed tasks and subtasks, and at the subtask level the displays and display elements were identified to which pilots would likely attend in order to obtain the information they required.

Also considered in this analysis of displays and tasks were the capabilities of the ACFS to support realistic manipulations of mental workload and the subset of these manipulations that should be selected for use in the planned validation studies. Likewise, alternative converging measures of workload were considered and the feasibility of obtaining various measures in the ACFS was delineated.

It became apparent during this analysis that it would be premature to attempt a comprehensive validation study of ERP indices of workload without first exploring, in a more limited way, the reasonableness of our conjectures regarding candidate displays and the extent to which pilots would be expected to observe and attend to these displays during realistic scenarios. Therefore, a need was identified for pre-validation behavioral and, if possible, occulometric testing in the ACFS, as a prelude to ERP recordings. Such testing is needed in order to confirm that we are focusing on the display events which command the pilot's attention at crucial times during a scenario.

3.3 Findings from the Task Analysis of ACFS Displays and Aircrew Tasks

Three ACFS displays proved most useful for our purposes. These three displays, the primary Flight Display with attitude format, the Navigational Display, and the Engine Power Display, are presented in Addendum 3A-1. Below, we list some changes in elements of these displays, and a few other cockpit events that appear to be candidates for eliciting ERPs related to workload. The approach and landing scenario that was adopted for the analysis of likely pilot use of displays is summarized in Addendum 3A-2. Possibilities for manipulating mental workload in the ACFS are listed below, as are a number of possible converging measures that could be examined in conjunction with ERP changes.

3.3.1 Candidate Display Events

The ACFS at Langley was running at a 16 Hz display update rate at the time of the present analysis, and it was slated to eventually run with a 20 Hz update rate. With these rapid cycle times, the displays change smoothly and continuously, without the "jumpiness" that would result from a slower update rate. It is not feasible, therefore, to time-lock ERPs to the updating of the displays per se. The present analyses did, however, reveal a number of candidate display elements which change discretely in task situations that would be expected to meet the criteria for eliciting endogenous ERP components related to mental workload:

o   The onset of various discrete messages on the Flight Display (see Addendum 3A-1), namely:

OUTER MARKER

MIDDLE MARKER

INNER MARKER

DECISION HEIGHT

FLARE BAR

FLAPS?

STEER POINT

Upon initiation, these messages flash several times. The ERP of interest would be elicited by the initial onset of each message. These ERPs may reflect the extent to which the pilot is preoccupied with other mental processing.

o   Changes in the color of the bar graphs on the Engine Power Display, reflecting the movement of engine performance parameters into Caution or Warning zones.

o   The appearance or disappearance of the Indicated Air Speed Deviation bar on the Flight Display.

o   The appearance or disappearance of the Radar Altitude Digital Readout on the Flight Display.

Other display changes pose an interesting possibility for eliciting endogenous ERP components. This category of "events" is not related to a discrete onset, offset or abrupt perturbation in a display, but rather to a smoothly changing display element attaining a state or screen position that is meaningful to the pilot. It is not clear whether such changes will elicit an ERP at all, but it should be recalled that endogenous ERP components such as P300 can be emitted, as opposed to evoked, in the absence of external stimulation (e.g., Ruchkin, Sutton, & Stega, 1980) if the information conveyed by that absence is meaningful to the subject. If preliminary empirical studies suggest that the following display changes elicit such ERPs, it is likely that these responses will be modulated by attentional demands:

o   When the position of the Flight Director Ball on the Flight Display leaves or enters the target area within which the pilot is trying to maintain it.

o   When the Air Speed "bug" on the Flight Display exceeds a specified level.

o   When the moving bars on the Engine Power Display reach or fall below thresholds which signify Caution or Warning conditions.

o   When the changing digital values for Altitude, Indicated Air Speed, Radar Altitude, Vertical Velocity, and the To and From Waypoint Altitudes on the Flight Display match or depart from commanded or intended levels.

o   Likewise, when the analog-like dials and indicators for Altitude, Air Speed, Flight Path Angle, and Horizontal Deviation and Track Angle Error on the Flight Display match or depart from commanded or intended levels.

o   When the Lubber Line on the Navigational Display crosses the Track Marker or Aircraft Heading Index.

Finally, other cockpit events, which are not CRT-based, but which meet our criteria for eliciting ERPs, should be mentioned. These include:

o   Onset of Caution and Warning indicators.

o   Onset of auditory annunciators to alert the aircrew to emergency conditions.   There are three types of auditory signals, reflecting three levels of problem severity.

3.3.2 Possibilities for Manipulating Workload in the ACFS

Although the aircraft modeled by the ACFS will contain a number of "intelligent" capabilities and semi-automated systems designed to lower the workload of the aircrew, the power and flexibility of this simulator provides many opportunities to vary workload, attentional demands and task difficulty. For example:

o   Several display formats (e.g. the Navigational Display) offer cluttered or decluttered alternatives.

o   Cross-winds, turbulence, and wind shears are modeled and can be introduced into different phases of flight.

o   Engine failure is modeled and can be used to introduce additional sensitivity into the "feel" of the controls.   This manipulation can be introduced in gradations of difficulty — e.g., no loss, loss of engine with center line thrust, loss of engine with asymmetrical thrust.

o   Various alarm conditions requiring an immediate pilot response (e.g., engine fire) can be introduced.

o   The sensitivity of the sidearm controller can be varied, giving the control of the simulated aircraft a different "feel".

o   Different airports are modeled and flight path approaches of varying difficulty can be manipulated.

o   The Flight Director Ball, a compelling pilot aid, can be switched on or off.

o   The frequency of occurrence or memory demands entailed by verbal

communications with ground-based air traffic controllers can be varied in realistic ways.

o   The frequency of required navigational changes enroute, for example to avoid weather patterns, can be varied.

3.3.3 Converging Measures of Workload

A number of possibilities exist for implementing other measures of workload in the ACFS for use in validating and evaluating the relative sensitivity of ERP effects:

o   Subjective ratings such as the SWAT technique (e.g., Reid, Shingledecker, & Eggemeier, 1981) or the NASA TLX technique (e.g., Hart & Staveland, 1986) are readily implemented. Granted, these measures entail either intruding on the pilot's task performance, in order to make the ratings on-line, or deferring the ratings until after the scenario, in which case crucial information can be lost. Nonetheless, these measures are easy to obtain and can prove enlightening.

o   The ACFS provides for the logging of many measures, derived from flight performance, that may vary with task difficulty and workload. The pilot's adherence to an ideal flight path or landing pattern should be revealing. Speed and appropriateness of response to emergencies or unexpected occurrences may be useful, although it may be necessary to instruct the pilot to respond rapidly if such measures are intended. It was noted that few conditions in the ACFS require a rapid response and many pilots are of the opinion that it is better to do nothing until the appropriate response is apparent rather than to act with undue speed and rashness. Despite the aforementioned limitations of contrived secondary task performance measures, if naturalistic secondary tasks can be defined, the speed and accuracy of performing such tasks should be useful.

o   Other physiological measures may also be instructive as concomitants of task difficulty and mental workload, although in many cases their validity is as much in question as that of ERPs. Eye blink rate, duration and

latency have been related to workload (e.g., Bauer, Goldstein, & Stern, 1987), and are particularly useful under relatively low vigilance conditions. Heart rate and heart rate variability have likewise been shown to be sensitive to workload manipulations, at least under certain conditions (e.g., Roscoe, 1982; Veldman, Mulder, Mulder, & van der Heide, 1985). Voice stress measures such as those presently being examined at NASA Langley, if they prove reliable, could be readily gathered from verbal communications by the pilot with ground control.

3.4 <u>Discussion and Recommendations for Studies to Validate the Results of These Analyses</u>

A general conclusion that can be drawn from the present analyses is that there are events in modern cockpits and cockpit simulators that are candidates for eliciting ERPs related to mental workload. Moreover, the general methods and issues examined in our analyses proved to be workable and should generalize to other settings.

Clearly, the present conjectures about display changes that will elicit ERPs of interest need to be validated. Likewise, the ease with which workload can be manipulated and converging measures that can be obtained in the ACFS need to be tested. Ideally, such validation testing should take place in several, well-considered steps. First, systematic "dry runs" without ERP recording should be conducted. The purpose of these studies would be twofold — to determine the likelihood that pilots will observe the display elements on which the candidate events occur, when they occur, and to confirm that the workload manipulations to be used in the subsequent ERP studies are salient enough to influence the more conventional measures of workload against which ERP susceptibility is to be judged. If problems occur in the initial aspects of validation, they will need to be addressed by further analyses or ACFS experimentation before the possibility of ERP indices of workload in the ACFS can be subjected to a fair evaluation.

These "dry run" tests should involve the manipulations of workload that are planned for the subsequent ERP studies. The introduction of cross-winds or turbulence while landing, or the failure of an engine upon final approach

appear to be good candidates for manipulating aircrew workload. The extent to which these manipulations affect subjective, behavioral and flight performance measures of workload should be examined. Pilot's attention to displays should be assessed by having the pilot "think aloud" as he performs the scenario and, if possible, by monitoring eye fixations and scan patterns using an oculometer. If it is problematic that pilots will observe the candidate display changes as they initially occur, it may be advisable to also use the oculometer during the ERP-recording studies to sort out the trials on which workload-related ERP changes would be precluded by inappropriate (for the present purposes) eye fixation or scan patterns.

It should be recognized that the present analyses were conducted with the constraint of identifying cockpit events that lend themselves to conventional ERP recordings. The power of working in the simulator setting could be more fully realized by some advances in ERP recording and analysis techniques:

o   An ability to trigger ERPs off eye movements themselves would greatly increase the number of candidate events one could consider and the number of trials one could expect to obtain in a given scenario. The brain's response to the display information provided when the eye fixates on a new display element after a saccadic movement, regardless of whether the external display changed or not, may elicit a similar response to that which occurs when the display itself changes. If so, these ERPs will also likely be modulated by workload. This possibility seems to be readily testable.

o   One could further expand the number of candidate events if it were possible to derive the measures related to workload on the basis of single trial ERPs. The present project, of course, included a task that was aimed at further developing such capabilities.

o   As mentioned previously, it is possible that certain continuous display changes would elicit an ERP when the changing indicator passes some point or enters some condition that is significant to the pilot. If empirical tests of this phenomenon should suggest the feasibility of "time-locking" to such continuous changes, the number of display events available for

analysis would be greatly expanded.

Some further observations about future directions and the degree to which the present ACFS findings generalize are in order:

o   While the ACFS provides many advantages as a test-bed for the transition of ERP measures from the laboratory to more practical settings, it also provides some impediments. Because the simulation is updated at 16-20 Hz, the displays appear to change more continuously than would be the case in less powerful simulators. Therefore, in many less high-fidelity simulators, the "jerkiness" in displays could be used to advantage in that, with each update cycle of the simulation, a perceptually discrete stimulus is provided to which ERPs could be time-locked.

o   Simulators with an out-the-window visual scene would create additional possibilities for introducing workload manipulations. Task difficulty could be readily manipulated by day versus night flying comparisons or by varying the cloudiness of the visual scene.

o   Cockpit Display of Traffic (CDTI) formats were not considered here in any detail, because none are planned for installation in the Langley ACFS. However, preliminary analyses suggest that this display format, and the situations in which it would be used, contain a number of ERP-eliciting events that might be related to workload. Interestingly, some aspects of these displays appear to be conceptually analogous to the task we designed for the empirical laboratory work conducted under the present contract.

o   Although the present analyses focused on high workload conditions, lower than optimal mental workload may be more of a problem in sophisticated, semi-automated systems such as the ACFS. If the pilot is lulled into a false sense of security or loses vigilance, he may not be prepared to respond appropriately when uniquely human capabilities are called into play by emergencies or other unexpected conditions. ERPs and other physiological measures may be sensitive indicators of such pilot mental states, so, at some point, analyses should be performed and studies conducted to examine the problems associated with low levels of workload.

Of course, under such conditions, there may be, by definition, fewer cockpit events being attended by the aircrew, so more ongoing indices of physiological activity (e.g., EEG, heart rate variability, EOG) may prove more fruitful than ERPs.

ADDENDA TO CHAPTER 3

ADDENDUM 3A-1

KEY DISPLAY FORMATS FROM ADVANCED CONCEPTS FLIGHT SIMULATOR

These display formats are taken from:

Sexton, G. A. (1983). Crew Systems and Flight Station Concepts for a 1995 Transport Aircraft (Report No. 166068). Hampton, Va.: NASA Langley Research Center.

ADDENDUM 3A-1.1 -- FLIGHT DISPLAY WITH ATTITUDE FORMAT (Figure 22)

ADDENDUM 3A-1.2 -- NAVIGATION DISPLAY FORMAT (Figure 24)

ADDENDUM 3A-1.3 -- ENGINE POWER DISPLAY (Figure 29)

ADDENDUM 3A-1.4 -- SAMPLE FLIGHT DISPLAY WITH SUPERIMPOSED MESSAGES

Figure 22. Flight Display with Attitude Format, (Sheet 1 of 4)

#	SYMBOL	COLOR	LOCATION & MOVEMENT
1	AIRCRAFT & CIRCLE	White	Center of ADI ball. Always stationary. Shows aircraft attitude in reference to horizon line and pitch and roll scales. Max climb or descent is 90°. Max roll is 90°.
2	HORIZON LINE	White	Across the ADI ball. In relation to aircraft symbol it moves up during descent, down during climb, and pivots around center of aircraft symbol to indicate roll. Left end of horizon line is lower during right bank and higher during left bank. Area inside the ball above the horizon line is shaded cyan; below the horizon line, black.
3	PITCH SCALE	White	Centered above and below horizon line. Remains parallel to horizon line during roll maneuvers. Shows 20° above and below aircraft symbol.
4	ROLL SCALE	White	Centered above pitch scale. Rotates with the horizon line as the aircraft banks. Scale markers are at 10, 20, 30, 45, and 60 degrees. Angle of bank is shown under the roll index. Max angle of bank is 90°.
5	ROLL INDEX	White	Centered on outside of ADI ball at the top. Always stationary. Shows angle of bank by pointing to the roll scale.
6	DELETED		

#	SYMBOL	COLOR	LOCATION & MOVEMENT
7	FLIGHT DIRECTOR BALL	Amber (Filled)	Moves left, right, up or down to indicate commanded vertical and lateral track. When the circle is flown so as to encircle the flight director ball, the aircraft will be flying the correct vertical and lateral profile to intercept or remain on desired paths. Disappears when flight director fails or is turned OFF
8	IAS DEVIATION BAR	Amber (Filled)	Bar grows down from bottom of left wing of the aircraft symbol to show deviation below commanded indicated airspeed (IAS). It grows up from top of the left wing to show deviation above commanded IAS. The length of the bar changes at a rate of ½ inch per 10 KIAS deviation to a maximum of 1.9 inches (38 KIAS). It disappears completely with plus or minus 2 knots deviation.
9	IAS SCALE	White	Occupies fixed position in upper left corner of display Scaled in 5 knot increments from actual aircraft IAS (shown in digits and under the tip of the pointer). It shows a range of plus or minus 50 knots from actual IAS. The numbers on the scale change at a point 180° from the tip of the pointer. The total range of airspeed is from 0 to 999 knots.
10	AIRSPEED DIGITAL READOUT & LABEL	White	Upper center of airspeed circle. Digits show the IAS of the aircraft and agree with the position of the pointer on the airspeed scale. When acceleration or deceleration is so rapid that the last digit changes too fast to be readable, only the even numbers are displayed. If it gets too fast again only the 0s and 5s are displayed. Total range is from 0 to 999 knots.
11	AIRSPEED POINTER	White	Extends from center to edge of airspeed scale, pivoting around the center. Tip points to the actual IAS on the scale.
12	COMMANDED AIRSPEED INDEX	Amber	Moves around the circumference of the airspeed scale and points to the commanded indicated airspeed. Difference between this index and the airspeed pointer is shown on IAS Deviation bar. The index disappears from view when the commanded value is more than 50 knots from the indicated.
13	MACH DIGITAL READOUT & LABEL	White	Lower center of the airspeed circle. Digits show the Mach of the aircraft. It has a range from 0.40 to 1.0 Mach. It disappears from view below 0.40.
14	BAROMETRIC ALTITUDE SCALE	White	Occupies fixed position in upper right corner of the display. Scaled in 100 foot increments indicated by single digits from 0 to 9.

Figure 22.   Flight Display with Attitude Format, (Sheet 2 of 4)

#	SYMBOL	COLOR	LOCATION & MOVEMENT
15	BAROMETRIC ALTITUDE DIGITAL READOUT	White	Upper center of altimeter circle. Digits show the barometric altitude of the aircraft. The last three digits reflect the position of the tip of the pointer. The total range necessary is from 0 to 50,000 feet. Between 1000 and 9990 the first digit is replaced with a hatched box. Below 990 feet only the 1, 2, or 3 digits are displayed, without hatched boxes. The number is rounded to the nearest 10 feet so that the last digit is always 0.
16	BAROMETRIC ALTITUDE POINTER	White	Extends from center to edge of altimeter scale, pivoting around the center. Tip points to the 100 feet of altitude scale.
17	HORIZONTAL DEVIATION & TRACK ANGLE ERROR SCALE	White	Centered below ball in a fixed position. Used with horizontal deviation pointer above and track angle error pointer below.
18	HORIZONTAL DEVIATION POINTER	Box-White Letters –Green or Amber	Moves along the top of the scale. Indicates amount of lateral deviation from flight plan. It is a "fly-to" indicator similar to the CDI on an HSI. Full scale is plus or minus 3 nm while enroute or 1 nm on approach when aircraft is outside the final approach fix (FAF). When inside the FAF and within 2.5° horizontal deviation, full scale changes to 2.5°. Pointer stops moving when deviation reaches full scale at which time the letters change from green to amber. When inside the FAF and within 2.5° horizontal deviation, letters change from H to L.
19	TRACK ANGLE ERROR POINTER	Amber	Moves along the bottom of the scale to indicate TAE of aircraft track from desired course. Full scale is plus or minus 20°. Pointer stops moving at max deviation. When aircraft track and desired course are parallel, regardless of whether it is on desired course, TAE pointer is centered. It moves away from center in the opposite direction to which the aircraft track (nose of aircraft with no wind) moves from a position parallel to desired course.
20	VERTICAL DEVIATION SCALE	White	Centered along right-hand side of ADI ball in fixed position.
21	VERTICAL DEVIATION POINTER	Box-White	Moves along the vertical scale to indicate vertical deviation from flight plan. Full scale is plus or minus 500 feet when aircraft is outside the FAF. When inside the FAF and within 2.5° TAE, full scale changes to plus or minus 0.7° vertical deviation from glide slope. Pointer stops moving when deviation reaches full scale at which time the letters change from green to amber. When inside the FAF the letters change from V to G.

Figure 22. Flight Display with Attitude Format, (Sheet 3 of 4)

#	SYMBOL	COLOR	LOCATION & MOVEMENT
22	RADAR ALTITUDE DIGITAL READOUT	Amber	Right center of display. Indicates height above ground below 2500 feet. Disappears above 2500 feet AGL. R indicates radar altitude. Readout shows lower than 200 feet - nearest foot; above 200 feet - nearest 10 feet.
23	VERTICAL VELOCITY POINTER & DIGITAL READOUT	White	Right center of display. Indicates descent with a down-arrow or climb with an up-arrow. Digital readout indicates feet per minute of change. Readout shows less than 500 ft/min, nearest 10 foot/min; 500 to 2000 ft/min, nearest 50 feet/min; above 2000 ft/min, nearest 100 ft/min.
24	TO WAYPOINT	Green	Lower right corner of display. Indicates desired altitude of the TO waypoint (the one the aircraft is proceeding towards) in hundreds of feet. Changes to indicate the altitude of the next waypoint when over or 90° abeam the TO waypoint.
25	FROM WAYPOINT ALTITUDE	Green	Lower right corner of display. Indicates flight planned altitude of the waypoint that the aircraft has just passed (FROM waypoint). Changes when the TO waypoint changes.
26	ANGLE-OF-ATTACK	Red (S) Green (O) Amber (F)	Left center of display. Split donut indicates on proper angle-of-attack (AOA) for weight and configuration. "F" indicates too fast or too low AOA. "S" indicates too slow or too high AOA. As AOA changes one-half the donut fades away, as the other symbol comes into view. F or S gets more pronounced as donut completely disappears.
27	FLIGHT PHASE & MODE ANNOUNCIATION	Green	Lower left corner of display. Indicates selected mode of flight and status of that selection.
28	BAROMETRIC PRESSURE	Green	Lower right corner of display. When barometric set knob is pulled out, altimeter setting is displayed in milibars and inches of mercury. Digits disappear when knob is pushed in.
29	FLIGHT PATH ANGLE INDEX	Amber	Right center of display. Moves along flight path angle scale.
30	FLIGHT PATH ANGLE SCALE	Amber	Right center of display. FPA scale is always centered on the aircraft symbol and has a range of ±6° FPA.
31	CARDINAL HEADING MARKERS	Green	Under horizon line. Move laterally across horizon line to show each 10° of track change.

Figure 22. Flight Display with Attitude Format, (Sheet 4 of 4)

Figure 24. Navigation Display Format, (Sheet 1 of 3)
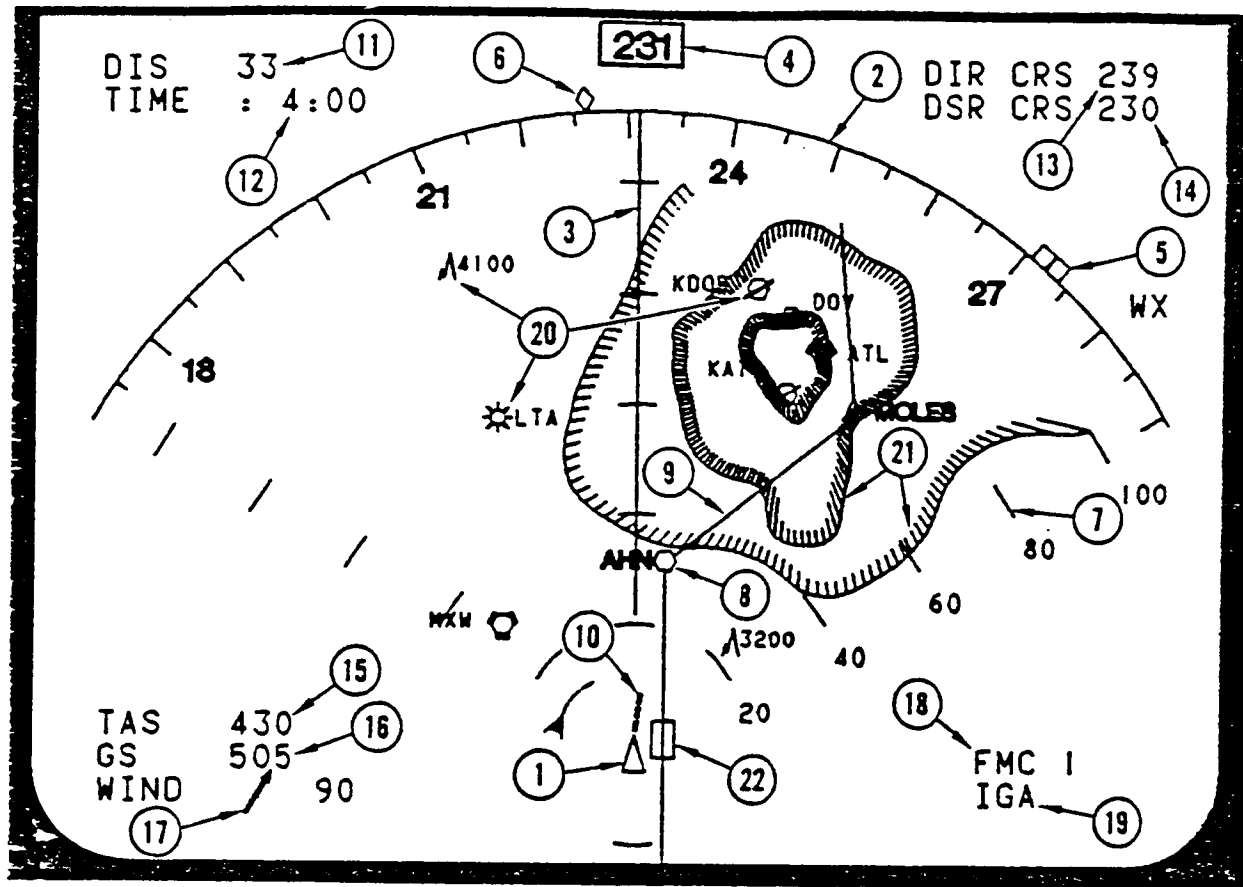
#	SYMBOL	COLOR	LOCATION & MOVEMENT
1	OWN AIRCRAFT	White	Always remains in a fixed position with the uppermost point of the triangle at the center of the smallest range marker circle.
2	COMPASS ROSE	White	120° arc with aircraft track at top center. Arc is divided into 5° increments with digits each 30° starting at 0°. Arc position does not move but scale changes as aircraft turns.
3	TRACK LUBBER LINE	White	Always oriented vertically from the own aircraft symbol to the compass rose. Shows aircraft track on the scale.
4	AIRCRAFT TRACK DIGITAL READOUT AND BOX	White	Above top center of compass rose. Shows numbers from 001 to 360 degrees and agrees with reading under lubber line on compass rose.
5	TRACK MARKER	Amber	Rotates around circumference of compass rose. It is positioned by pilot.

#	SYMBOL	COLOR	LOCATION & MOVEMENT
6	AIRCRAFT HEADING INDEX	Amber	Rotates around circumference of compass rose. Shows aircraft heading on the compass rose and drift correction angle as the difference between aircraft heading index and track lubber line.
7	RANGE MARKERS	White	Equidistant marks between Own aircraft and maximum range selected. Marks maintain fixed position. Scaled in NM and identified on right hand end.
8	WAYPOINT SYMBOLS AND IDENTIFIERS	White	Three dimensional points defining route of flight. They move with respect to Own aircraft symbol at rate based upon ground speed and range scale selected. When aircraft is on desired track, nearest waypoint is always shown vertically above Own aircraft symbol.
9	PLANNED COURSE LINE	Green	Line between waypoints defining route of flight. Moves with the waypoints. Terminates at largest range marker or furtherest waypoint on one end and bottom of display on other end.
10	POSITION PREDICTOR OR TREND VECTOR	White	Three dashed lines extending from front tip of Own aircraft symbol. The end of each dash shows the predicted position of the tip of the Own aircraft symbol at 20, 40 and 60 seconds from the present time based upon present aircraft course and ground speed. Lines change length with respect to ground speed and display range scale.
11	DISTANCE TO GO DIGITS	Green	Upper left corner of display. Shows nautical miles between tip of Own aircraft and nearest (TO) waypoint. Figure is shown in full miles (no decimal) until under 10 NM; then miles and tenths of miles.
12	TIME TO GO DIGITS	Green	Upper left corner of display. Shows hours, minutes and seconds required to travel from present position to next (TO) waypoint. Leading zeros (insignificant) are not shown.
13	DIRECT COURSE DIGITS	Green	Upper right corner of display. Shows course (typically magnetic) between Own aircraft and next (TO) waypoint.
14	DESIRED COURSE DIGITS	Green	Upper right corner of display. Shows course (typically magnetic) between last (FROM) waypoint and next (TO) waypoint.
15	TRUE AIRSPEED DIGITS	Green	Lower left corner of display. Shows true airspeed of aircraft.
16	GROUND SPEED DIGITS	Green	Lower left corner of display. Shows ground speed of aircraft.

Figure 24. Navigation Display Format, (Sheet 2 of 3)

#	SYMBOL	COLOR	LOCATION & MOVEMENT
17	WIND ARROW AND DIGITS	Green	Lower left corner of display. Shows wind vector (arrow) pointing from the direction that the wind is blowing relative to the aircraft track. Arrow disappears when wind is calm. Digits show wind velocity in knots.
18	NAVIGATION MODE	Green	Lower right corner of display. Shows navigation mode selected for display.
19	NAVIGATION SOURCE		Lower right corner of display. Shows sources of navigation signals being used to obtain navigation display.
20	NAV AIDS, AIRPORTS AND OBSTACLES	Green Amber	Symbols for nav aids, airports and/or obstacles may be selected for display.
21	WEATHER RADAR CONTOURS	Red Green Amber	Weather radar returns may be selected as an overlay to the map.
22	TIME BOX	Amber	Appears when TNAV is selected on GCP. Moves along desired course line of moving map. Indicates the position that the aircraft should be in to arrive at a metering fix at a particular time.

Figure 24. Navigation Display Format, (Sheet 3 of 3)
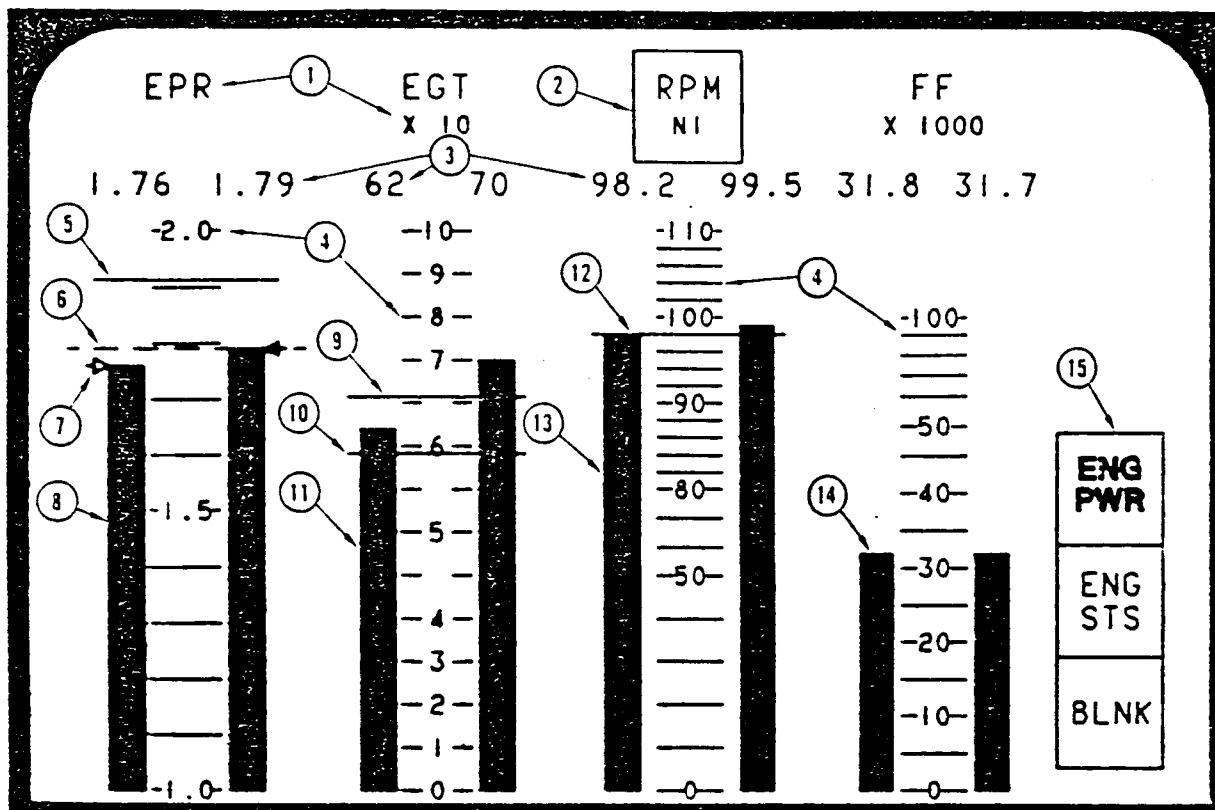
1. ALL PARAMETER LABELS AND UNITS - GREEN

2. RPM TOUCH PANEL SWITCH OUTLINE - GREEN

3. DIGITAL READOUT OF PARAMETERS -

   - EPR - NORMAL, GREEN; WARNING, RED

     EGT - NORMAL, GREEN; CAUTION, AMBER; WARNING, RED

     RPM - NORMAL, GREEN; WARNING, RED

     FF - GREEN

4. ALL SCALES - WHITE

5. EPR LIMIT MARKER - RED

6. COMMANDED EPR MARKER - WHITE

7. THROTTLE EPR SETTINGS - AMBER

8. ACTUAL ENGINE EPR BARS - NORMAL, GREEN; WHEN THE EPR LIMIT IS EXCEEDED, THAT PORTION OF THE BAR ABOVE THE LIMIT LINE PLUS THE INNER HALF OF THE VERTICAL BAR BELOW THE LIMIT LINE TURNS RED

9. EGT WARNING LIMIT MARKER - RED

10. EGT CAUTION LIMIT MARKER - AMBER

Figure 29. Engine Power Display, (Sheet 1 of 2)

11. ACTUAL EGT BARS - NORMAL, GREEN. WHEN EGT EXCEEDS THE CAUTION LIMIT, THAT PORTION OF THE BAR ABOVE THE CAUTION LIMIT LINE PLUS THE INNER TWO-THIRDS OF THE VERTICAL BAR BELOW THE CAUTION LIMIT LINE TURNS AMBER. WHEN EGT EXCEEDS THE WARNING LIMIT PLUS THE INNER THIRD OF THE VERTICAL BAR BELOW THE CAUTION LIMIT TURNS RED; THE OUTER TWO-THIRDS OF THE BAR BETWEEN THE WARNING AND CAUTION LIMITS PLUS THE CENTER THIRD OF THE VERTICAL BAR BELOW THE CAUTION LIMIT REMAIN AMBER; AND THE OUTER THIRD OF THE VERTICAL BAR BELOW THE CAUTION LIMIT REMAINS GREEN.

12. RPM LIMIT MARKER - RED

13. ACTUAL RPM BARS - NORMAL, GREEN; WHEN RPM LIMIT IS EXCEEDED, THAT PORTION OF THE BAR ABOVE THE LIMIT LINE PLUS THE INNER HALF OF THE VERTICAL BAR BELOW THE LIMIT LINE TURN RED

14. ACTUAL FUEL FLOW BARS - GREEN

15. TOUCH PANEL SWITCHES - GREEN

16. SWITCH LEGENDS - WHITE IF SELECTED (SHOWN AS HEAVY FONT), GREEN OTHERWISE

Figure 29. Engine Power Display, (Sheet 2 of 2)

Engine Power Format - The engine power (ENG PWR) format displays engine pressure ratio (EPR), exhaust gas temperature (EGT), either N1 RPM or N2 RPM selectable with the touch switch around the label, and fuel flow (FF).

FLAPS
DESCEND MAX 2000

20

10

RWY 10000

HOLD
TURB
1 ENG

DES
ALT CHP
ALT ARM
ATT CSS
LOC ARM
TCMD

M 90 MAX

M 90 MAX

OM MM IM
ALERT

0000R

TO: 000
FM: 000

000000
00.00

050

TIME : 4:00
DIS 11.0

DIR CPS 050
DSP CPS 050

36

33   03

30   06
                                   200

KLMNO   FGHIJ

                        160

          ABCDE   120

                  80

THS 193   40   FMC 1
GS 190   MLS 1
WIND 000   TRUE

PQRST

          0000   0000
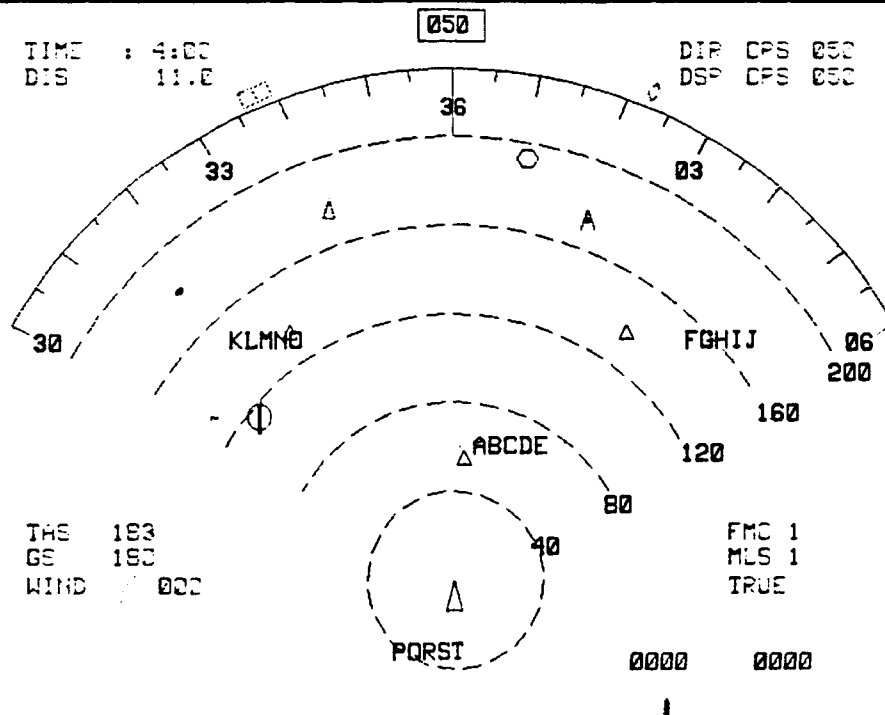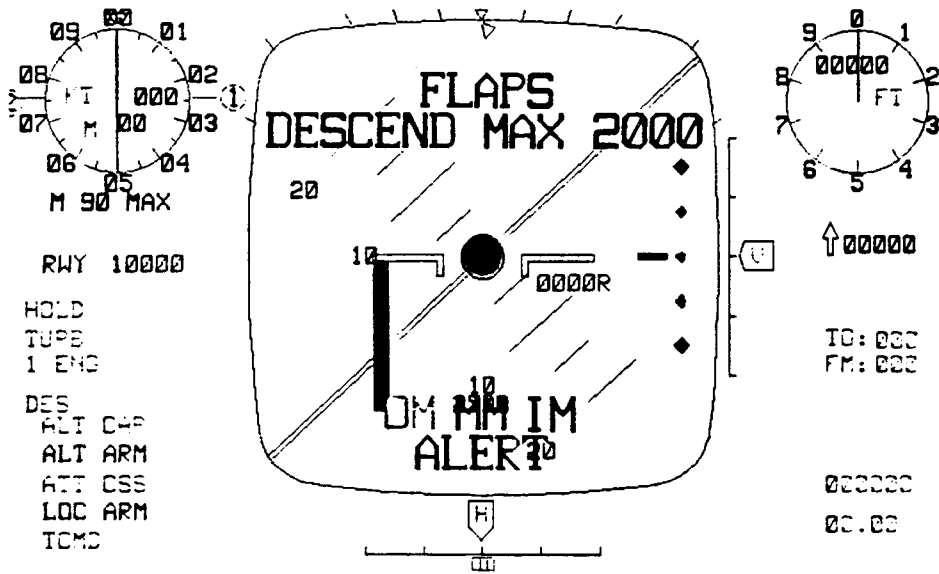
SAMPLE FLIGHT DISPLAY WITH SUPERIMPOSED MESSAGES

ADDENDUM 3A-2

SUMMARY OF REPRESENTATIVE LANDING SCENARIO

FOR WORKLOAD ANALYSIS IN THE ADVANCED CONCEPTS FLIGHT SIMULATOR

The following table indicates flight parameters and control actions that are likely to be of primary interest to the ACFS aircrew during various stages of a representative approach and landing. The scenario is broken down in terms of tasks (columns) and the aspects of flight that are being controlled (rows):

PITCH = Pitch Axis control

ROLL = Roll Axis control

THRUST = Thrust Axis control

ANT = Anticipatory control of the aircraft

NAV = Navigational control of the aircraft

DISCRETE CONTROL = Control of the aircraft involving individually distinct movements or mediating responses

The flight parameters and control actions in the cells of the table are coded as follows:

ADD = Increase throttle

ADJ = Adjust throttle

ALT = Altitude

ANGLE OF BANK (Self explanatory)

AS = Air Speed

DIST = Distance

FLAPS (Self explanatory)

FLT ANG = Flight Angle

GEAR (Self explanatory)

GS = Ground Speed

POSIT = Position

RANGE MARKERS (Self explanatory)

RET = Retard throttle

TIME (Self explanatory)

TRK = Track (i.e. Flight Path)

WIND (Self explanatory)

	ENROUTE TO HOLDING I				HOLDING II			
	ENTRY (A)	DECREASE AIR SPEED (B)	MAINTAIN AIR SPEED (B)	MAINT CRUISE (C)	COMMENCE & MAINTAIN TURN (A)	ROLLOUT & MAINTAIN LEVEL LEG (B)	CONTINUE TURNS (C)	CONTINUE LEVEL LEGS (C)
PITCH	ALT AS	ALT	ALT AS	ALT AS	ALT AS	ALT AS	ALT AS	ALT AS
ROLL	TRK	TRK	TRK	TRK	ANGLE OF BANK	TRK	ANGLE OF BANK	TRK
THRUST	ADJ	RET	ADD	ADJ	ADD ADJ	RET ADJ	ADD ADJ	RET ADJ
ANT		POSIT	AS		POSIT	TRK	POSIT	TRK
NAV	POSIT TIME DIST WIND GS			POSIT TIME DIST WIND GS	POSIT	POSIT TIME DIST WIND GS	POSIT	POSIT TIME DIST WIND GS
DISCRETE CONTROL	RANGE MARKERS	SET AS						

	ARCING III	TRANSITION TO FINAL APPROACH IV				GLIDE SLOPE V	
	COMMENCE & MAINTAIN TURN	ROLLOUT & MAINTAIN LEG (A)	DECREASE AIR SPEED (B)	MAINTAIN AIR SPEED (B)	MAINTAIN LEG (C)	COMMENCE & MAINTAIN DESCENT (A)	LAND DECISION (B)
PITCH	ALT AS	ALT AS	ALT	ALT AS	ALT AS	AS FLT ANG	
ROLL	ANGLE OF BANK TRK	TRK	TRK	TRK	TRK	TRK	
THRUST	ADD ADJ	RET ADJ	RET	ADD	ADJ	RET ADJ	
ANT	POSIT	TRK	POSIT	AS		POSIT	ALT
NAV	POSIT	POSIT TIME DIST WIND GS			POSIT TIME DIST WIND GS		
DISCRETE CONTROL			SET AS GEAR FLAPS			FLAPS	

3-29

4.0 DATA ANALYSIS SOFTWARE UPGRADE

As part of the present effort, ARD upgraded an existing, proprietary data analysis software package. This software was used for analysis of ERP data from the present laboratory experiments and a version of it was delivered to NASA. This software was initially written by Dr. Daniel S. Ruchkin and his assistants at the University of Maryland School of Medicine to support an extensive range of ERP data analysis capabilities. ARD obtained the rights to this software from Dr. Ruchkin in order to support the analysis of data from the present project as well as others. We named the package ANALYZ and upgraded it in a number of ways, as detailed below.

There were some limitations, for the present purposes, in the initial implementation of the software, due to the fact that it was developed to run on a PDP-11/40 computer under the RT-11 Single Job operating system (RT-11SJ). This environment is one in which available memory is limited. The RT-11SJ monitor and associated device handlers use approximately 6K 16-bit words of memory (of the total available 32K), thus requiring that programs occupy no more than 22K words. As a result, the ANALYZ programs were heavily overlaid and array sizes were limited. These constraints entailed some severe limitations when these programs were applied to data sets from repeated measures designs involving numerous experimental factors with several levels of each factor. In addition, certain routines were coded in PDP-11 assembly language, decreasing the package's portability to other systems. Finally, the graphic display capabilities of the package were limited and the graphics routines were specific to the VT-11 display with which the package was initially used.

As part of the present effort, ARD converted the ANALYZ package to run on ARD's VAX-11/730 computer under the VMS operating system. In so doing, a number of features of the code were changed to enhance the portability of the package. This resulted in a more powerful version of the package for both RT-11 and VMS systems, and one which can more readily be ported to other systems in the future. In addition, the existing graphics routines were rewritten and the

package's interactive graphics capabilities were enhanced. The graphics routines were structured in a modular fashion, such that most of the data management and user-system interaction code were implemented in generic FORTRAN routines, with the calls that are specific to the particular color graphics terminal that we used (a Raster Technologies One-10) being embedded in separate FORTRAN subroutines. With this configuration, the programs could be ported to another system with only the latter routines needing modification. In the following sections we offer an overview of the capabilities of ANALYZ and a description of the upgrades that were accomplished under the present contract.

4.1 Overview of ANALYZ

4.1.1 Summary of Capabilities

The ANALYZ software first operates upon a disk file of single-trial ERPs, where a "trial" refers to the response(s) elicited by one occurrence of a particular event. At the beginning of each such data file is a header record which contains an alphanumeric label and the numeric parameters that are common to the entire file (e.g., the number of channels, inter-sampling interval, number of time points in the analysis epoch, etc.). Following the header record there is a record for each ERP trial. Each ERP record consists of:

o   "Identifying" parameters, which describe such characteristics as:

   --   Stimulus and response conditions under which that trial was elicited.

   --   Timing information, such as when in a sequence of stimuli this trial was elicited.

   --   Code for subject's behavioral response.

   --   Flags to indicate condition information to the experimenter.

o   Information about each scalp channel, such as:

   --   Root-mean square (RMS) variability of the digitized time points.

   --   Mean pre-stimulus baseline levels.

   --   Number of time points that were "clipped" by the A/D converter.

o   The ERP waveforms for each channel.

The programs that operate upon single-trial data feature the following operations:

o   Inspection — display and plotting of single-trial ERP waveforms, with optional digital filtering.

o   Editing — correct or modify parameters, correct measurements of baseline and/or RMS levels.

o   Selection and summarization — indicate which trials have parameters which meet specific conditions, compute statistics for specified parameters, RMS levels and/or baseline levels.

o   Computation of average waveforms — averages are computed over trials that meet specified conditions. The averages can be synchronized to either a stimulus or to a response. A program is available that will compute latency-corrected averages via the Woody latency correction algorithm (Woody, 1967).

Average ERPs are stored in files consisting of a header record, which indicates the current number of average ERPs in the file and a 72-character comment that contains information that is common to the whole file of ERPs (e.g., the subject and experiment), followed by a record for each ERP in the file. Key information in each average ERP record consists of:

o   Alphanumeric labelling information — a 36-character comment that is specific to an individual average ERP (e.g., the electrode location and experimental condition).

o   Numeric labels (e.g., the experiment number, subject number).

o   Identifying parameters (e.g., time at which the stimulus occurred, number of time points in the analysis epoch, number of time points in the pre-stimulus baseline, inter-sample time).

o   The average ERP waveform.

o   The associated standard deviation waveform.

o   If relevant, reaction time or latency statistics (including histograms).

There are a number of programs that then operate upon the average ERPs. Programs of this type are concerned with operations such as the following:

o   Inspection — display and plotting of average ERP waveforms, with optional digital filtering.

o   Grand averaging — pooling across average ERPs to compute grand average ERPs within-subjects or across-subjects.

o   Computing difference waveforms.

o   Extraction of amplitude measures — baseline-to-peak or area under the curve measures, with optional digital filtering.

o   Data management — e.g. listing of average ERP files, moving ERP records between files.

o   Principal Components Analysis — computation of component loading waveforms for a set of average ERPs, with optional digital filtering; computation of component scores as an estimate of ERP component amplitude.

The derived measures of ERP amplitude may then be subjected to several types of further analysis, including:

o   Inspection — display and plotting of average ERP amplitudes, with the data nested.

o   Analysis of Variance — repeated measures ANOVA with Geisser-Greenhouse or Epsilon correction for Type I error.

o   Multivariate Analysis of Variance — repeated measures MANOVA with optional "profile analysis."

Data and control parameters are stored in files with standardized formats. Two kinds of files were alluded to above, namely single-trial ERP files and average ERP files. The same format is used for files of unpooled average ERPs, files of ERPs pooled across conditions and/or subjects, and files of average ERP difference waveforms. In order to speed up the execution of programs that operate via sorting on parameters such as experimental conditions, condensed disk files can be generated which contain parameters used for sorting, but do not contain waveform data. These condensed files are used for high-speed look-ups and, when data (e.g., single-trials) are found that match the specified sorting conditions, the original, complete data file is then accessed. In addition to the above-mentioned data files, there are also data files which contain amplitude measures (e.g., base-to-peak at a given time point or average area under the curve) extracted from average ERP waveforms.

Depending upon its type, a program may operate in either an interactive or a batch mode. Some programs can operate in both modes. Most programs have specialized control parameter disk files associated with them. These files contain information that directs the operation of the program. In some cases, such as in an interactive display program, the control file may be relatively small. In such applications it serves as a means of "remembering" specific control parameters, so that the user need not type in control parameters each time the program is invoked. If the control parameters are satisfactory, then the program can immediately start its operation (e.g., display). In programs that run in a batch mode the control parameter file may be relatively large (e.g., for computing average ERPs from single-trial data, the specifications for computing each average ERP are contained in the control file). Typing such large amounts of information directly into a computer at the start of program execution can be a cumbersome, error-prone procedure. Thus there are programs that prepare such files before executing the batch run. Such programs take advantage of nesting arrangements in the data to "semi-automatically" build up the control file and provide rapid listing routines so that the contents of the control files can be readily checked.

4.1.2 Summary of Data File Types

The programs prompt the user for the file names from which to input data or

control information. Thus the files can be named by whatever nomenclature the user chooses. However, there is considerable advantage in adopting a convention for naming certain kinds of files. The following convention is assumed here and by the menus, prompts, and other messages presented to the user on the screen:

.SSD — Single-trial data file.

.TTD — Condensed single-trial data file, containing trial-by-trial identification parameters but not the digital waveforms themselves.

.PAR — File containing control parameters for computing average ERP waveforms.

.OUT — Average ERP waveform file.

.WOD — Latency-corrected average ERP waveform file.

.ACS — File of average ERP waveforms pooled across subjects; same structure as .OUT file.

.TTS — Condensed average waveform file (used for merging).

.DFF — File of difference ERP waveforms

.BPR — Extension used for files that contain control parameters for a given program. The structure of these control files depends upon the program — in effect, the disk "memory" for program control parameters. This extension is usually used for batch mode operations.

.001 — Numeric extensions, preceded by the program name, are also used to designate control parameter files for a given program. This kind of file designation is usually used for the disk "memory" of interactive mode operations.

.BTP — File containing baseline-to-peak amplitude measures extracted from .OUT type files.

4.1.3 Summary of Main Programs

The ANALYZ package should be viewed as a set of software tools for understanding particular sets of ERP data. The programs employed and the options chosen will depend on the data and the experimental design within which it was collected. Nonetheless, it is easiest to summarize the package's capabilities by examining the programs in the sequence with which one would

typically explore such a data set:

General utility program:

CRFIL1 — "Create file" — create files for data or control information; copy subsets of data from one file to another.

Main programs associated with inspection, editing, and displaying single-trial ERP data:

EDSSD1 — "Edit SSD files" — inspect/edit files of single-trial ERP data.

EDTTD1 — "Edit TTD files" — apply auxiliary identifying information, i.e., that which was not coded with the ERP data at the time of data collection, to files of single-trial ERP data; summarize single-trial data based on identifying information.

DISSD1 — "Display SSD files" — display single-trial ERP waveforms.

Main programs associated with sorting and averaging the single-trial data and displaying average ERP data:

EDPAR1 — "Edit PAR files" — construct control files for selective averaging.

SOSSD1 — "Sort SSD files" — sort single-trial data to be selectively averaged, using a compressed file of identifying information only to obtain a quick sort.

AVSSD1 — "Average SSD files" — average single-trial ERPs time-locked to eliciting stimuli or behavioral response times.

PWOOD1 — "Prepare for Woody" — set up for Woody latency-correction analysis.

WOODY1 — "Woody analysis" — average single-trial ERPs with latency

adjustment for trial-to-trial latency "jitter."

DIOUT1 -- "Display OUT files" -- display average ERPs.

Main programs associated with data management of average ERP data, extracting derived measurements from average ERPs, and displaying these derived measures:

LIOUT1 -- "List OUT files" -- list average ERP data files and identifying information.

EDBTP1 -- "Edit baseline-to-peak" -- derive amplitude measures from average ERP waveforms.

DIBTP1 -- "Display baseline-to-peak" -- display amplitude measures.

DIFWV1 -- "Difference waves" -- compute difference waveforms between pairs of average waveform data.

MERGE1 -- "Merge" -- merge average waveforms within or across subjects.

4.2 Upgrades for Transportability

The tasks which ARD performed in upgrading the ANALYZ software were aimed at producing a package of programs and subroutines which could be easily transported to any environment which supports FORTRAN. The environment into which the software was initially ported was the VMS operating system (Version 3.4) using the VAX VMS (Version 3.4-56) compiler. Specifically, the following tasks were performed:

o   Replacement of RT-11 System Specifics -- Virtually all of the I/O functions (e.g., file opening/closing, reading/writing from/to disk) in the RT-11 version of ANALYZ were accomplished using system service routines and/or MACRO-11 assembly language routines which were called from FORTRAN subroutines. A library of VAX FORTRAN routines, which emulated the functions performed by this combination of system service calls and assembly-language subroutines, was written in VAX FORTRAN. The goal in

constructing these routines was to provide the programmer with a calling form for file opening and closing, and for input/output which was consistent with those used in the RT-11 version of ANALYZ. The VAX-specific routines were named differently, however, to insure that the RT-11 versions of these FORTRAN-callable subroutines were replaced in all routines.

o   DCL Compilation and Linking Routines -- The indirect command files (.COM files) which were used under RT-11 to direct the compilation and linking of programs were converted to their functional equivalent in the VAX/VMS Digital Command Language (DCL). In addition, these .COM files were modified to allow for the relatively automatic updating of object modules when changes to source code had been made. One important difference between the RT-11 and VMS versions of the .COM files concerns overlay structures. Recall that under RT-11 there is very little available memory space (approximately 44K bytes) to accommodate code and data. As a result, in many programs in the RT-11 version of ANALYZ, it was necessary to "overlay" object modules such that only those modules which referenced each other were simultaneously memory-resident. Under VMS, such memory restrictions are largely transparent to the applications programmer and are handled by the task swapper.

o   Conversion of all Macro-11 Routines to FORTRAN -- In addition to the file I/O functions discussed above, there were many assembly language routines used in the RT-11 version of ANALYZ for other purposes. These assembly language routines had been implemented primarily in order to speed up numerical calculations. While it was possible to convert these Macro-11 routines to their equivalent in VAX-11 assembly language, it was decided that the goal of portability of the package took precedence over the relatively small gain in computational speed that would result from using assembly language instead of FORTRAN. As a result, VAX FORTRAN routines equivalent to those implemented in Macro-11 assembly language were written. As always, the interface to the calling routine was maintained.

o   Identification of DEC-Specific Conventions -- With some relatively minor exceptions, VAX FORTRAN 77 provides for compatibility with RT-11 FORTRAN in

their extensions to the ANSI X3.9-1978 standard of FORTRAN 77. In the course of converting from RT-11 FORTRAN to VAX FORTRAN 77, we identified a number of features which, although not requiring a change at this time, might cause compatibility problems if the code were ported to another system. Some examples of such features, which may require attention at some point in the future, are: 1) storage of characters in LOGICAL*1 variables; 2) use of ENCODE and DECODE statements to convert between numeric and character variables; 3) use of the $ character to suppress carriage returns; and 4) use of DATA statements for initialization of variables in COMMON.

4.3 Graphics Enhancements

It was necessary to develop an enhanced graphics capability for the display of ERP data since the graphics capabilities of the initial RT-11 version of ANALYZ were largely specific to display devices which are obsolete (i.e. the VT-11 display processor for CRT graphics and a Houston Instruments plotter). In addition, there was some inflexibility in the design of the display programs in the RT-11 version of ANALYZ and, because they had been written for a monochrome display, they didn't exploit the power of color displays for revealing patterns in the data. In order to address the need for a display capability, ARD implemented a series of FORTRAN callable subroutines for both CRT displays and hardcopy plots of the following:

o   Single-trial ERPS.

o   ERP average waveforms.

o   Baseline-to-peak amplitude measures.

o   Principal Components Analysis weighting coefficients and basis waveforms.

The display terminal selected for CRT display of data was the Raster Technologies Model One/10. This terminal has 640 x 480 x 8 display memory, 60 Hz refresh rate, a 255-color lookup table, and is typical of the type of relatively inexpensive display terminal used throughout the field. It is a

bit-mapped device which uses display list architecture, such that when an element of the display is changed, only the state of those pixels affected by the change are updated; the balance of the display need not be redrawn into display memory.

Raster Technologies supplies a library of FORTRAN-callable subroutines with the One/10 ("ONELIB"), thus obviating the need for development of primitive functions.   During Phase II, ARD defined the functions necessary for interactive examination of ERP data and designed and implemented VAX FORTRAN programs, using the ONELIB library, to perform these functions.   The features built into these programs were ones which allowed the experimenter to dynamically configure the display on the basis of the characteristics of his data. These features included:

o   Up to 16 waveforms/line drawings per set of axes.

o   User-specified color and line type of each waveform.

o   User-specified horizontal and vertical scales.

o   Option for different vertical scales for different sets of axes.

o   User-specified polarity of waveforms.

o   Cursor movement and digital readout of latency and amplitude.

o   Optional FIR filtering of waveforms.

o   Displays of ERP waveforms, standard deviation waveforms, or response time
     histograms.

In addition to this CRT graphics display capability, ARD also implemented a series of FORTRAN programs which allowed for hardcopy plots which were analogous to the graphics displays.   These hardcopy plots were implemented using a Hewlett-Packard 7550A plotter and HPGL command set.   FORTRAN routines were written which allowed for more mnemonic calling conventions for the basic

functions necessary to construct these plots. These routines were then called by FORTRAN programs to generate the resulting plots. In addition to the features implemented in the CRT graphics display programs, the hardcopy plotting programs also implemented additional features which allowed for the flexible of near publication quality plots. These features included:

o   User-specified line type and color.

o   User-specified number of rows, columns, and waves/lines per page.

o   Legends and header labels.

o   Automatic checking of user layout to determine if appropriate records should be plotted on the same axes.

5.0 DEVELOPMENT OF AN APPROACH TO STUDY ALTERNATIVE METHODS
OF ERP SINGLE-TRIAL ANALYSIS

5.1 Background and Overview of the Single-Trial Analysis Task

Scalp-recorded ERPs are usually extracted from the ongoing EEG by signal averaging the brain activity recorded in response to numerous occurrences of an eliciting stimulus. The waveform that is recorded in response to a single such event is referred to here as a "single-trial." The number of trials typically included in time-locked averages range from several dozen to several thousand, depending upon the type of ERP activity of interest and its signal-to-noise ratio with respect to the background EEG. For ERP components related to cognition (e.g., see Donchin, et al., 1978), several dozen trials are typically averaged.

For many applications in operational settings, such signal averaging is impractical, because the eliciting conditions can change unpredictably and one wishes to quantify ERP activity on a moment-to-moment basis. Even if stimulus conditions can be presented repeatedly, as they can for some "open-loop" studies in which system design issues are being addressed and the data can be analyzed off-line, there is a problem with ERP components varying in latency from trial-to-trial. Because the components of interest are related to cognitive processes, and in complex tasks the timing of these processes may vary from trial-to-trial, the ERP components may "jitter" in time from one trial to the next. Average ERPs calculated under such conditions will contain broader, lower amplitude waveshapes than were present on the single-trials that were averaged together.

Several techniques have been developed to extract useful information from ERPs on a single-trial basis (see recent review by Childers, et al., 1987). Step-Wise Discriminant Analysis (SWDA) has proven to be remarkably successful in characterizing ERP amplitudes in single-trial responses (Donchin & Herning, 1975; Horst & Donchin, 1980; Squires & Donchin, 1976; Squires, et al., 1976). However, SWDA assumes that the latencies of ERP components will be fixed from

trial-to-trial. Other techniques have been proposed to quantify the trial-to-trial variability of ERPs (e.g., Coppola, et al., 1978), but again the implicit assumption has been that the ERP signal is relatively stable in time.

Dealing with latency variability in ERP components, either between or within conditions, has been more difficult. One technique that has been used is a cross-correlational approach developed by Woody (1967). A template that corresponds to the ERP waveshape of interest is cross-correlated with a segment of EEG in which that waveshape occurs, and the offset at which the maximal correlation occurs is chosen as the latency of the component. This technique has been used by a variety of ERP investigators over the last ten to fifteen years, but a number of known problems in its application have become apparent.

Aunon and McGillem (1975) have proposed an alternative technique which, in some respects, is different. It precedes the cross-correlation with a minimum mean square error (MMSE) filter (i.e., Wiener filter) and has provisions for independently latency adjusting different regions of an ERP. This technique has not been as widely used as the Woody filter, so its virtues and limitations are not as well known.

A systematic comparison of these two techniques on simulated data has not been attempted. Such an analysis would be a major step towards clarifying the conditions under which latency varying ERPs can be identified and quantified on a single-trial basis. The purpose of the present task was initially to conduct such a comparative analysis and, in the course of so doing, to develop the necessary software to create simulated EEG records with known ERPs embedded in them.

Early on in this project it became apparent that a direct comparison between these two techniques would not be straightforward. As discussed in Chapter 2, both techniques contain a number of user-selectable parameters and options, and the extent to which an appropriate set of choices is made largely determines how well each technique will fare under a given set of conditions. When dealing with real data, one obviously has much less information about the signals that are present in a waveform than when dealing with simulated data. Nevertheless, in practice, the sophisticated data analyst usually has some idea

of the characteristics of the signals that are present. Therefore, we found the need for an exploratory software tool that would allow various analyses to be implemented interactively, with differing assumptions being made about the nature of the underlying data.

Another initial impediment was that our existing software implementations of the Woody and Aunon/McGillem approaches were both stand-alone packages of routines that were somewhat cumbersome to use. Both packages were written in FORTRAN, but the Woody routines were implemented on ARD's VAX-11/730 system, while the Aunon/McGillem routines were implemented initially on a PC-AT and modified to run on the VAX. There was little commonality in the input formats required, the statistics of merit calculated, or the output measures made available. In the VAX environment, graphics output was available only in post-processing display and hard-copy plotting routines.

Therefore, we focused our efforts on creating a software environment to support the sort of exploratory analyses that seemed desirable. As discussed more fully in Section 5.3, we followed an integrative approach, implementing aspects of the Woody and Aunon/McGillem techniques within the framework of a program which also generates the simulated EEG/ERP data, calculates common statistics of merit, and provides an on-line display of the output of various stages of the analysis process. The numerous user-selectable parameters are input to the main program from ASCII control files which can be modified using any standard text editor. The analyses are implemented in a manner which supports user interaction and encourages exploratory data analysis. At the same time, provisions were made for setting up batch process "production runs" using a data management "shell" program. This feature allows the user to specify a series of comparative analyses, systematically varying a user-selectable parameter or set of parameters, for unattended execution.

Source code for the software developed here has been delivered to NASA along with user documentation. The results of initial comparative analyses conducted with this software are presented in Section 5.4. These analyses confirmed some of our preconceptions about the strengths and weaknesses of the Woody and Aunon/McGillem approaches and revealed aspects of the two approaches which had not been apparent from published reports. No far-reaching conclusions about

the relative virtues of the two techniques seem warranted as yet, but it is apparent that the software approach developed here can provide the basis for a variety of future explorations of the conditions under which the Woody and Aunon/McGillem techniques may be useful. Some future directions for the use and/or enhancement of this software by either ARD or NASA are presented in Section 5.5.

5.2 Discussion of Alternative Latency Correction Techniques

5.2.1 Overview of the Woody Approach

C. D. Woody (1967) reported on the development of a generic, adaptive technique for extracting a latency-varying signal from background noise. He applied the technique to a variety of neuroelectric data recorded from anesthetized cats. The technique worked successfully for detecting electrical stimulation responses and auditory evoked potentials recorded from the surface of the cortex, eye blink potentials elicited by mechanical stimulation of the glabella, and miniature synaptic potentials recorded from single motor units in the spinal cord. It was also applied to the detection of petit-mal spike-and-dome discharges in human scalp recordings. An attractive feature of the technique appeared to be that, with sufficient signal-to-noise ratios, little or no prior knowledge was required about the waveshape of the signal. Additional properties of this approach for functioning under different signal-to-noise conditions were described by Harris and Woody (1969).

Figure 5-1 presents the flow of this data analysis approach. The heart of the technique involves a cross-correlation between a template of the signal of interest with single-trial recordings in which that signal is embedded, in a latency-varying manner. For each such single-trial record, the template is cross-correlated at a variety of time lags (offsets) and the lag at which the cross-correlation is maximal is taken to be the latency of the signal in that record. Over trials, a latency-corrected average of the signal is constructed by aligning the single-trial records time-locked, not to the time of occurrence of the eliciting stimulus, but to the point at which each record was maximally correlated with the template.
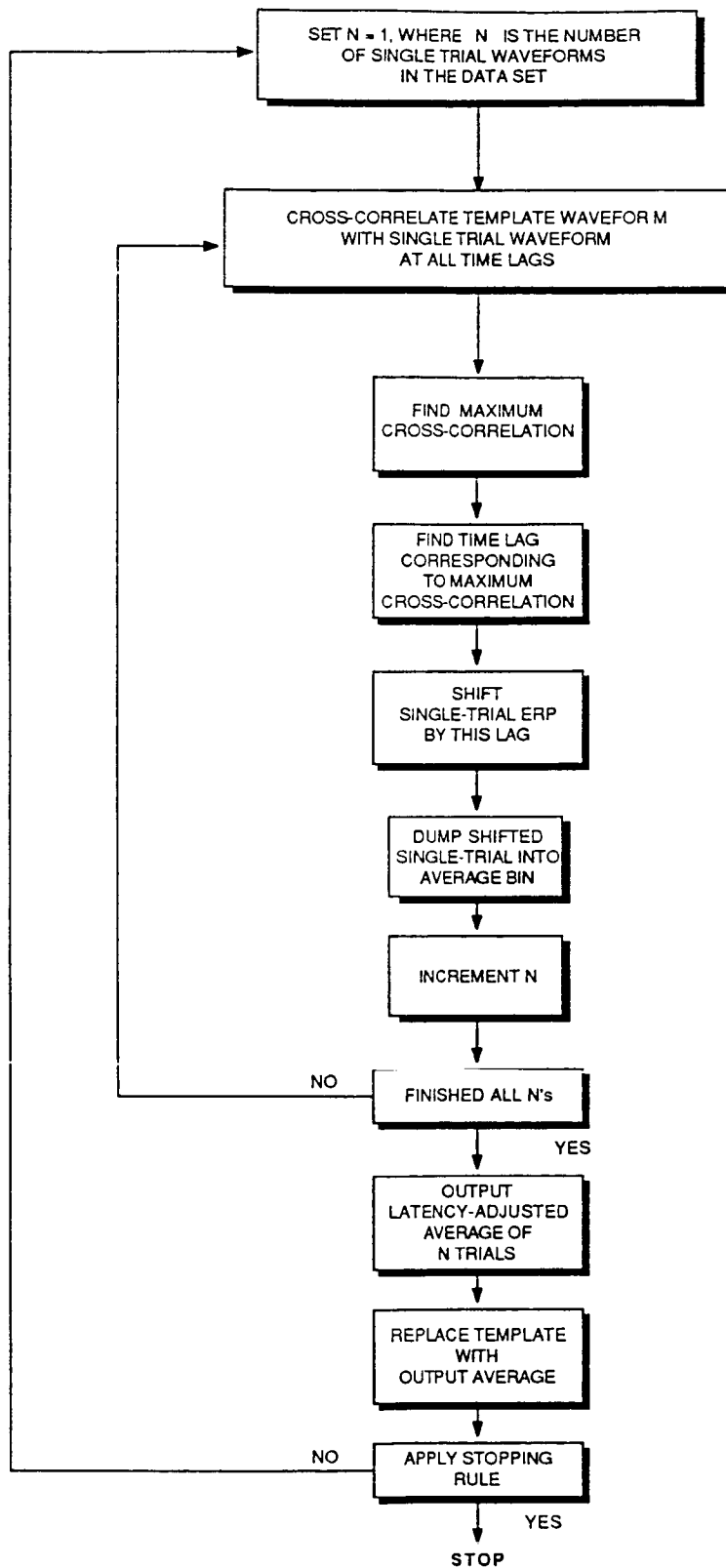
Figure 5-1. Flow of the Woody Analysis (Adapted from Woody, 1967)

The adaptive nature of the algorithm stems from the choice of the template waveform. The latency-corrected average from one pass through a given set of data can be used as the template for the next pass. With iterations through the data, this technique successively approximates the waveshape of the underlying signal. The stimulus-locked average is typically used as the template for the first pass. Various stopping rules, such as the failure of the mean maximal cross-correlation (of template to single-trials) to improve by a specified amount, can be invoked to determine when to end this iterative process.

The application of the Woody approach for latency-correcting human ERP data was pioneered by Ruchkin (see Ruchkin & Sutton, 1978, 1979; Ruchkin, et al, 1980, 1981) and has been employed by a number of other investigators (e.g., see Kutas, et al., 1977; Horst, et al., 1980; Wastell, 1977 for some of the earlier uses of the technique). These uses of the Woody technique proper had been preceded by several related cross-correlational analyses (e.g., Weinberg & Cooper, 1972; Pfurtscheller & Cooper, 1975). Typically, the ERP component of interest in studies utilizing the Woody technique has been P300. Data analysis issues have included characterizing the latency-corrected waveshape, amplitude and mean latency of P300 across different experimental conditions in which variables expected to affect amplitude and/or latency were manipulated.

Some serious problems with applying the Woody approach to scalp-recorded ERP data have become apparent, and, despite widespread recognition of the latency variability problem in interpreting endogenous ERP components, the approach has never gained general acceptance in the field. The problems in applying the Woody procedure concern the following issues:

o   The procedure will always result in a latency-corrected average having a sharper peak than was present in the raw, stimulus-locked average, even when the technique is applied to EEG data containing no ERP signal. Therefore, it is incumbent on the investigator to demonstrate that the latencies chosen by the cross-correlation across trials follow a non-random distribution. To address this concern, Ruchkin & Sutton (1978, 1979) have suggested applying Chi-square and Kolmogorov-Smirnov statistics to the histogram of latencies chosen by the technique. One can test the observed

histogram of latencies against the rectangular distribution that would be expected if latencies were being chosen at random, or against the latencies actually chosen when the technique is applied to "noise only" data (i.e. epochs of background EEG or a long-latency part of the ERP epoch beyond where the ERP peaks of interest would be expected to occur).

o   When ERP component amplitudes vary across experimental conditions, the latency estimates will be less reliable in the low amplitude conditions, and mean latencies derived for those conditions will tend towards the midpoint of the epoch that is being searched for the component of interest.

o   The technique is confounded when more than one component of the same polarity is active in the epoch searched. Two or more components shifting together from trial-to-trial would not pose a problem, but if more than one component "jitters" in time independently, the technique can be expected to pick one component on some trials and the other component(s) on other trials. In this latter case, the latency-corrected average will clearly be a misrepresentation of the underlying waveshape. Unfortunately, it is now clear that, under many task conditions of interest, there are typically multiple, positive-going endogenous positivities overlapping in the long-latency region of the ERP.

o   Although templates can be derived by an iterative process, whereby the latency-corrected average from one pass through the data is used as the template for the next pass, the fidelity with which the derived template corresponds to the underlying ERP component of interest is crucial. It is not yet clear under what signal-to-noise conditions the iterative process proposed by Woody will fail to converge on the underlying waveshape.

o   The signal-to-noise properties of the data, and thus the applicability of the Woody technique, can be greatly enhanced by pre-filtering the single-trial records. Most investigators, being interested in P300 and other relatively low frequency ERP components, have used a simple, low-pass, "box car" filter (see Ruchkin & Glaser, 1978). However, it is not clear to what extent the success of the technique is dependent on appropriate pre-filtering, and what, if any, advantages would be derived by

the application of other digital filtering techniques.

o   Woody (1967) identified some confounds that can arise in the pattern recognition process because of the nature of the cross-correlation algorithm.   Some practitioners have found a cross-covariance approach to be more satisfactory in that it better takes into account a correspondence between the amplitude of the template and the amplitude of waveshapes similar to the template that occur in the single-trial ERP. Alternatively, amplitude criteria could be used in conjunction with local maxima in the cross-correlation function to decide when an appropriate correspondence between template and single-trial waveshape has been found.

5.2.2 Overview of the Aunon/McGillem Approach

J. I. Aunon and C. D. McGillem (1975, 1979; also see McGillem & Aunon, 1976, 1977) developed an alternative pattern recognition approach which attempts to circumvent some of these problems with the Woody procedure. Their interest was specifically in the detection and latency correction of scalp-recorded human ERPs.   However, their studies focused on the ERPs elicited by simple visual stimuli in non-task situations, so the components of interest were the relatively high-frequency, sensory-related exogenous components of ERPs, rather than the slower frequency, longer latency endogenous activity that has concerned most practitioners of the Woody technique.

Figure 5-2 presents the flow of the Aunon/McGillem technique.   An aspect of this approach that has been given considerable emphasis by Aunon and McGillem is a pre-filtering of the single-trial data with a minimum mean square error (MMSE) filter (i.e. Wiener filter) in order to improve signal-to-noise ratios. After pre-filtering, the single-trial EEG record is cross-correlated with a template in order to identify peaks. This step is somewhat similar to the Woody cross-correlation, but differs in that Aunon and McGillem used an artificial waveshape (triangular wave) as their template, the process is not iterative (in the sense of the multi-pass Woody approach), and the goal is to identify all local maxima in the cross-correlation function, rather than the single latency at which the maximal cross-correlation occurs.   Over trials, a histogram is constructed of the latencies at which these putative peaks occur,
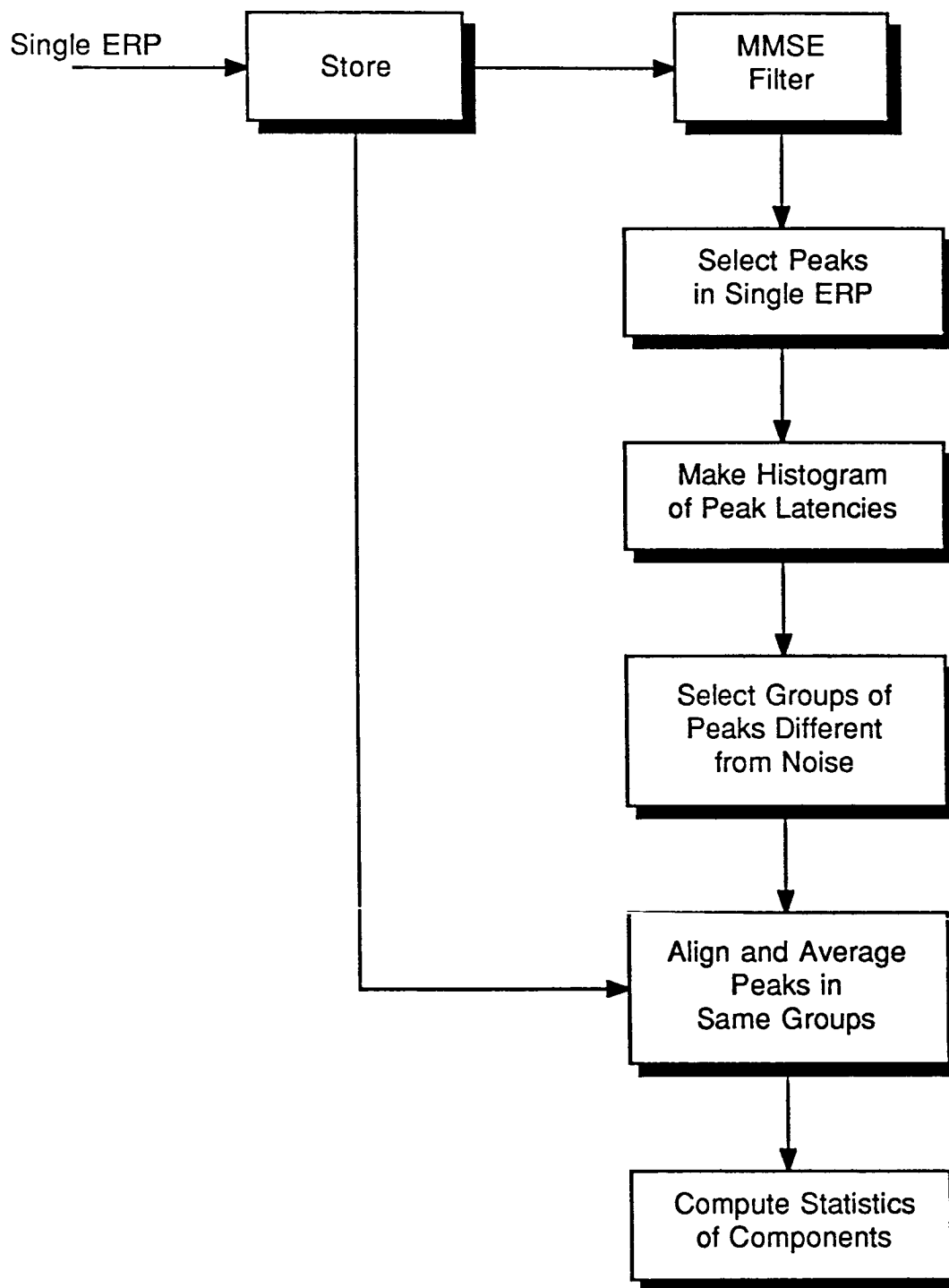
Figure 5-2. Flow of the Aunon/McGillem Analysis (adapted from Aunon & McGillem, 1979)

and statistical criteria are applied to the running mean of the histogram to identify those peaks which can be presumed to have come from a population that is not random in polarity, as noise would be. The running mean is also used to determine the regions over which peaks should be grouped together by combining into common groups those peaks lying between zero crossings of the running mean (Aunon & McGillem, 1979). Having thus identified the subepochs in which latency-varying peaks occur, the raw, unfiltered single-trial records are characterized with respect to the means and variances of the amplitudes and latencies of these peaks and with respect to the probabilities of occurrence of the various peaks. A latency-corrected average is then constructed by separately averaging time-shifted segments of the waveform in the subepochs delineating these peaks, and then piecing together these averages with certain assumptions being made to avoid discontinuities in the waveform.

Aunon & McGillem (1979) argue that their technique reveals latency variability in the visual ERP components which can not be explained by the effects of background EEG. The technique has been used by several other investigators (Aunon, personal communication), but with each customizing aspects of the algorithms to suit their needs. Therefore, the field as a whole has not had nearly the experience with the Aunon/McGillem technique as with the Woody technique.

The possible advantages of the Aunon/McGillem approach are that it provides a way of addressing the problem of multiple ERP components shifting independently in time from trial-to-trial and it offers criteria for inferring the presence or absence of a given ERP component on a particular single-trial, rather than assuming, as does the Woody approach, that the signal is always present. Nonetheless, many questions remain as to the signal-to-noise conditions under which this technique will be successful, the necessity of some of the steps in the peak-picking algorithm, and the criticality of the MMSE filter, as opposed to other filtering approaches, for the success of the technique.

5.2.3 User Choices and the Difficulty of Direct Comparisons Between Techniques

Clearly, systematic evaluations are needed of both the Woody and Aunon/McGillem

techniques applied to simulated data. However, as mentioned previously, it became apparent, as we familiarized ourselves with the inner workings of both techniques, that the outcome of these simulation tests could be easily biased for or against either technique by an informed choice of the many user-selectable parameters that are inherent in each approach. This fact highlights the extent to which the results obtained by an investigator working with real data will be influenced by the presuppositions that are made in implementing the analyses. Furthermore, it emphasizes the need for an exploratory approach to the analyses of simulated data, providing for a systematic evaluation of the accuracy of each technique given the fidelity with which the assumptions made by the investigator mirror the parameters which were used to generate the data.

The success of both techniques is critically affected by an appropriate choice of search epochs, i.e. the time points of an ERP in which the latency-varying waveform of interest is presumed to occur over trials. Choosing an epoch which fails to bound the ERP component(s) of interest will result in a distorted latency-corrected average, a truncated histogram of peak latencies, and incorrect mean latencies for the component of interest. Likewise, an appropriate choice of pre-filtering parameters will either enhance the signal-to-noise ratio of the component(s) of interest or confound the application of either technique by worsening the accuracy of the pattern recognition algorithms. The choice of an appropriate artificial waveshape as the template in the Aunon/McGillem procedure will obviously have a profound influence on the accuracy with which latency-varying peaks are identified.

Furthermore, as alluded to earlier, it is not clear to what extent either technique is dependent on some of the analytic choices made by its developers. At least under some conditions, the Woody procedure works better if a cross-covariance algorithm is used instead of cross-correlation. Might the same also be true of the Aunon/McGillem procedure? Would the Woody procedure benefit from a pre-filtering of the data by the Aunon/McGillem MMSE filter, or conversely, how well would the Aunon/McGillem approach work with the much less computationally intensive box-car filter? Could single-trial signal-to-noise criteria be added to the Woody procedure to provide an indication of trials on which no discernible signal is present? Might the Aunon/McGillem procedure benefit from a multi-pass, iterative approach modeled after the Woody adaptive

procedure?

These questions emphasize the fact that neither the Woody nor the Aunon/ McGillem technique, in its original form, may be optimal for application to human scalp-recorded data, as well as the fact that the two techniques need not be viewed as mutually exclusive. It may well be that an amalgamation of the two approaches, which utilizes some aspects of both, may prove optimal for dealing with ERP data.

These considerations, in addition to the logistical difficulties presented by the implementations of the two procedures which existed at the beginning of the project, led us to focus on the development of a software environment that would support the sort of exploratory analyses and further development of both techniques that seemed necessary.

5.3 Approach to a Comparative Evaluation of Alternative Techniques

5.3.1 Overview of a Comparative Approach

Because of the numerous interacting parameters that could affect the functioning of both the Woody and Aunon/McGillem techniques, the need to systematically explore the implications of user choices and assumptions on both techniques, and the desire to "cross-pollinate" the two techniques, a modular framework was adopted which integrated both analytical approaches. The software we developed was, therefore, designed to perform several functions:

o   To generate simulated EEG with user-selectable frequency parameters and root mean square variability.

o   To allow simulated ERP components (at present, one or two half sine-waves) to be embedded in the background EEG; the ERP components can be of a user-selectable amplitude and their latency "jitter" from trial-to-trial is governed by user-selectable latency parameters.

o   To implement parts of the Woody and Aunon/McGillem procedures for application to the simulated EEG/ERP records.

o   To provide as a "statistic of merit" the difference in latency (i.e. time
    point) between where the simulation algorithm positioned the ERP signal(s)
    on each trial and where each data analysis technique determined the signal
    to be.

The above functions were all integrated into a single program. This program
was written in the C language to run under the DOS operating system on an IBM
PC/XT/AT or compatible. In order to make this program a useful tool for
exploratory data analysis and to provide the user/analyst with a
straightforward, efficient means for modifying program parameters, it was
outfitted with a number of additional features:

o   Integrated EGA (enhanced graphics adapter) graphics which display
    cumulative records of several key parameters and waveforms on a
    trial-by-trial basis.

o   A control file for supplying key parameters to the main program; thus
    parameters can be changed by editing the control file, without the need to
    recompile the program each time a change is made. There are separate
    control files for data analysis and display parameters (see documentation
    supplied under separate cover).

o   Output statistics of merit are appended to the control file so that
    information about the program parameters which produced a given set of
    results are stored with the results.

The main program can run in either an "interactive mode," in which the user can
inspect the trial-by-trial graphics display for as long as he wishes before
moving on to the next trial with a carriage return, or in a "production mode,"
wherein the program runs to completion for a given set of user-defined input
parameters. In the production mode, the graphics display is still presented,
but it is updated at a rate which is pre-selected by the user and, therefore,
is not under interactive control.

It was anticipated that after initial exploratory data analysis in the
interactive mode, the user would want to define multiple sets of parameters and

execute a series of production runs, systematically varying one or more parameters of the data or data analysis process (e.g., signal-to-noise ratio of the ERP amplitude to background EEG; time epoch over which latency jitter could occur; search epoch over which data analysis algorithms operate). Therefore, a very flexible data management "shell" program was developed. This shell allows the user to readily define such a series of analyses and then execute the consequent production runs unattended, with results and a log being saved for later inspection. The shell routines are programmed in QuickBASIC*.

5.3.2 Data Simulation

The simulation of background EEG records is performed by feeding the output of a random noise generator, i.e. white noise, into a digital, autoregressive filter. The user specifies the center frequency, band width, and sampling time for this process. At the present time, up to two sets of frequencies may be specified, and in our preliminary studies we have used a combination of alpha and delta power bandwidths. The delta bandwidth, centered on a frequency of zero, is generated with a single-pole filter, while the alpha bandwidth, centered at approximately 20 Hz, requires a two-pole filter. Any given set of waveforms may be exactly duplicated by re-using the same initial seed given to the random number generator. During the initialization of the EEG generator routine, the root mean square of a sample of each component is recorded and used to normalize the subsequent output for each component to a preamplified power of approximately unity. The approach that we have implemented generally follows that developed by Zetterberg and associates (see Wennberg & Zetterberg, 1971; Zetterberg, 1969; Zetterberg & Ahlin, 1975; Narasimhan & Dutt, 1985).

A separate routine is used to simulate ERP signals. It can generate one or two half sine-waves of a specified amplitude and latency and add these waveshapes into the generated EEG record with the signal being randomly and independently jittered around its median latency. The probability spectrum for the jitter is a triangular shape, the base length of which is specified by the user. The probability envelope for the jitter may be displayed on the user's terminal.

---

*QuickBASIC is a trademark of the Microsoft Corporation.

The latency histograms resulting from the Woody and Aunon/McGillem procedures should, to the extent that either technique is successful, correspond in shape and position to these envelopes. As displayed, the relative amplitude of each triangle corresponds to the amplitudes of the generated signals.

A Fast-Fourier transform computes a power spectrum for the combined signal and background EEG. It is to be used primarily to provide a visual check on the above simulation techniques. It employs a simple Fast-Fourier algorithm, and it gives a meaningful spectrum only when the number of points of the waveform is an integral power of 2.

5.3.3 Partial Implementation of the Woody Technique

Unlike the Aunon/McGillem procedure, which uses a constant artificial template to correlate with the single-trial record, the Woody procedure creates a template based on the data itself. On the initial pass, a stimulus-locked average of the single-trial waveforms is used as the template, but subsequently the latency-corrected average computed from the previous pass is used as the template for the next pass. In the present implementation, the template evolves with each trial and can also be displayed with each trial. The user can watch as the template progressively assumes a shape closer to that of the input signal, and can decide at what point to halt the iterative process. The initial average template is created in a separate initialization process visually controlled by the user.

The template is subsequently correlated against the input waveform at every so many points as specified by the user. The position of the maximum point in this correlation function is taken to be the latency for the ERP signal for that trial. The latency is then recorded in the histogram of latencies and used to time-shift the input waveform before it is added to the running average waveform that will then serve as a template for the next trial.

5.3.4 Partial Implementation of the Aunon/McGillem Technique

Aunon and McGillem's implementation of their own approach embodies three main procedures — filtering, signal detection, and signal reconstruction. Only the

second of these procedures was implemented by the software developed here. By thus narrowing the focus of this study, we were able to make a more direct comparison between the basic detection schemes of the Woody and Aunon/McGillem approaches.

In the initialization of the Aunon/McGillem routine, a half sine-wave template of user-specified length is created. Unlike the Woody procedure, this routine uses a cross-correlation with the template only as a preliminary stage to the actual signal detection process, and so it will be tolerant to even a rough correspondence between template and signal.

In the main portion of the Aunon/McGillem routine, the template is correlated against the filtered input waveform at every so many points of the waveform, as specified by the user. The skipping of points will speed up this most time-consuming part of the procedure, but with a corresponding loss of resolution. The normalized correlation function may then be displayed.

Next the correlation function is passed to a peak detection routine. That routine checks for maxima by testing for changes in the sign of the slope. To be recorded, the putative peaks must also be greater in height than a specified fraction of the highest peak in the function. A cumulative count of the peaks is stored in a histogram. It is this histogram which is used to determine the placement of the signal detection "epochs." Within each of these epochs, a latency-corrected average waveform is recorded and displayed.

A single common length for each of the epochs is specified by the user. The middle of an epoch is placed at each maxima in the histogram of correlation peak frequencies, which meet certain criteria specified in the program and also by the user. Another routine creates and displays a histogram of the net number of positive minus negative peaks. It makes an absolute value copy of this histogram, similar to the Woody histogram of latencies, and passes it to another, quasi-recursive procedure. This procedure finds the maximum of the "latency" histogram and then determines which adjoining bins can be considered a part of that peak, i.e. the "shoulders", according to a combination of programmed and user-specified criteria. It then zeroes that peak and its corresponding shoulders. The above procedure is repeated on the remaining

parts of the histogram, until no part of the histogram is greater than a specified minimum, relative to the total number of peaks recorded thus far.

The program then determines if any of the correlation function peaks have occurred within a given epoch. If any of these peaks meet criterion, the routine will determine the appropriate latency correction and add a portion of the latency-corrected waveform to the average waveform being maintained within that epoch. After each trial, the accumulated average for each of the epochs may be displayed as a single, but discontinuous, wave.

5.3.5 Graphics Output

Figure 5-3 illustrates a typical display that would appear on the user's terminal for a run in which a single simulated ERP component was added to the simulated background EEG. All important parameters which control the data generation, application of analyses, and display are user-selected via control files.

The various frames displayed here are as follows:

o   EEG2 -- the simulated background EEG for the latest 1280 msec single-trial.

o   BOX -- the simulated single-trial EEG with simulated ERP added, with the resulting waveform having been digitally filtered with a box-car filter.

o   WOOD -- the running latency-corrected average that has been obtained thus far from application of the present partial implementation of the Woody procedure.

o   ENVL -- the triangular waveshape indicates the latency distribution of ERP center points with which the generated ERP components are being added to the background EEG records; the amplitude of this triangle is proportional to the amplitude of the simulated ERP waveshapes being generated; the histogram represents the latency histogram of points chosen by the Woody algorithm thus far; the RMSE value is the running root mean square latency error for the Woody procedure.

o   AUN -- the cross-correlation function that resulted from the cross-correlation of the artificial template with the latest single-trial waveform within the present partial implementation of the Aunon/McGillem procedure; the tick marks indicate the points in this function that have met the criterion for "peaks."

o   The unlabelled frame is the running histogram of peak latencies that have been accumulated across trials thus far by the Aunon procedure; the two
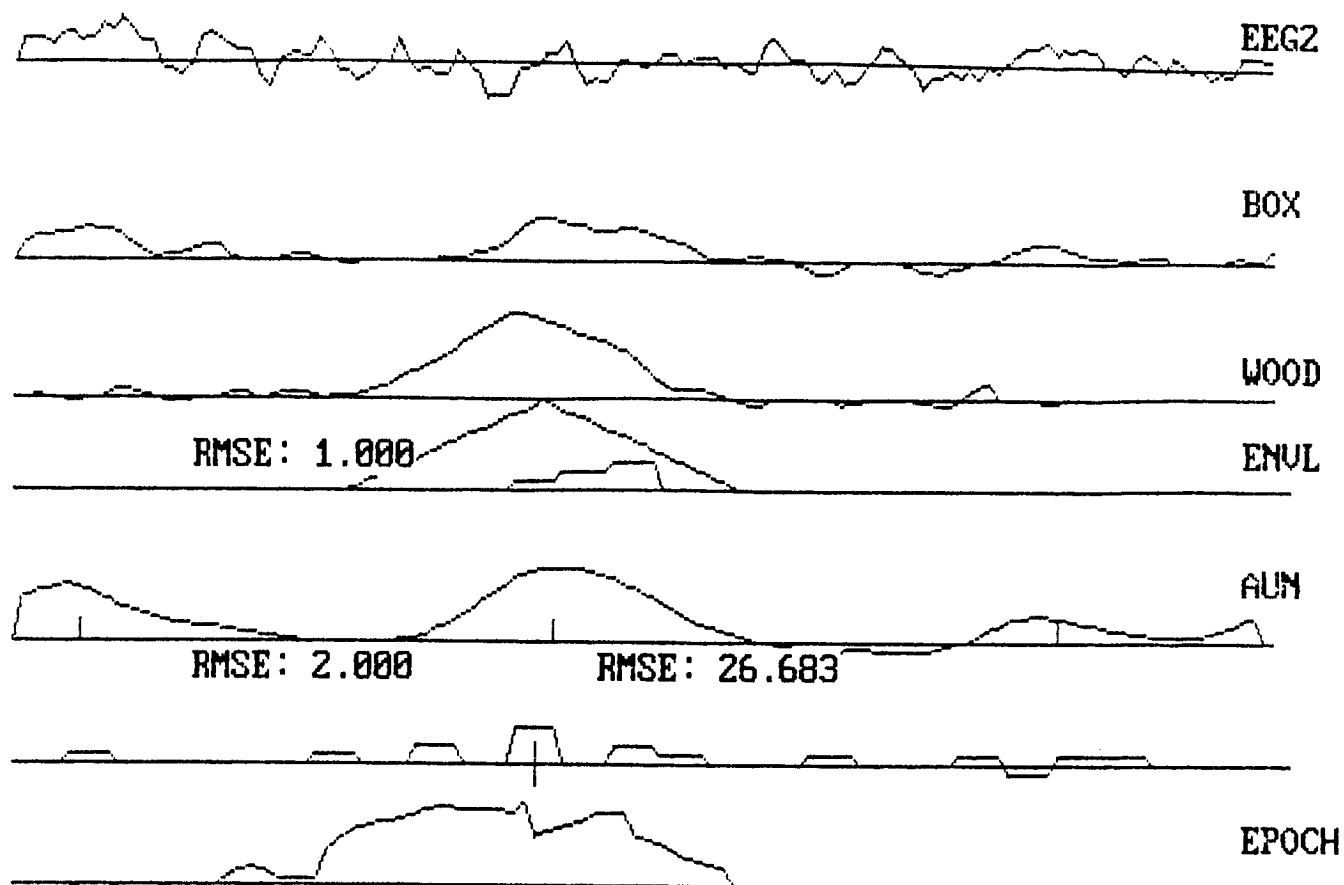
Figure 5-3. A sample display

RMSE values are the running root mean square latency errors for the Aunon procedure (note that because only one ERP component has been generated per trial for this run, the second RMSE value is not meaningful).

o   EPOCH — the running latency-corrected average that has been obtained thus far from the application of the present partial implementation of the Aunon/McGillem procedure.

5.3.6 Other Output

At the end of a data analysis run, the root mean square latency error for each procedure, in addition to being displayed on the user's screen, is appended to the control file of input parameters. This statistic of merit is calculated as the difference between the time point at which the generated ERP was actually centered and the time point at which the program determines that it was centered, averaged over trials. Mean latency errors are, of course, presented separately for the Woody and Aunon/McGillem techniques, providing a measure of

the relative performance of each technique.

5.3.7 Shell for Data Management of Production Runs

The data management shell provides for the automatic generation and unattended execution of a series of analysis "production" runs over which a specified input parameter(s) will be systematically manipulated. If multiple parameters are specified for manipulation, runs of the analysis routines will be generated for all combinations (nested) of the desired parameters. These routines accomplish the following objectives:

o   For every control file parameter, they allow the user to define the sequence of values that the variable is to take on over a series of production runs; changes in the values of more than one variable at a time can be specified.

o   Automatically generates, in a nested manner, the necessary series of appropriate control files with all combinations of the specified variables being varied (e.g., if ERP amplitude is to be varied from 1 to 7 units in steps of 2, and the period of the ERP half sine-wave is to be varied from 10 to 50 time points in steps of 20, the shell program will generate twelve control files, varying these two variables in a parametric manner).

o   Automatically calls the main data generation and data analysis program multiple times, passing it a different control file each time.

o   Traps errors so that, for example, inappropriate user choices of variable values in a given control file will not cause the whole production run to abort.

o   Keeps a log of this process so that the user has a quick reference for determining which analyses during a multiple-analysis production run executed successfully.

5.4 Results of Feasibility Tests

5.4.1 Functionality of the Software

The software developed here was shown to be quite functional. It provides the initiated user with a great deal of flexibility and control over the data generation and analysis process. On the one hand, the present implementation saves the investigator considerable time in exploratory data analysis by providing immediate, interactive feedback on the progress of a given analysis. Thus, analyses based on a particular combination of parameters can be initiated and, if found wanting, can be aborted, and another set of parameters can be chosen. On the other hand, the investigator who already has completed some exploratory, interactive analyses or who has a feel for what set of parameters will meet his needs, can readily specify a set of analyses to be run in the "production" mode. These analyses, which may run for hours, will then execute automatically, with results being stored conveniently along with the input parameters that generated them, and with a log of the production run being produced for later inspection.

The modularity of the present implementation has also proven useful in debugging code and adding features. Our initial experience, therefore, seems to confirm that the present implementation is workable and should foster its further evolution as a data analytical tool.

5.4.2 Results from Manipulations of Simulated Data

The results from initial analyses using this software are presented on the following pages. The following manipulations are illustrated:

o   Figures 5-4 through 5-6 -- Effect of varying the amplitude of the embedded ERP from a minimal signal-to-noise level to one which is easily detected by both techniques. At a poor signal-to-noise ratio (i.e., low amplitude ERPs, see Figure 5-4), the Aunon/McGillem approach provided a considerably more accurate estimate of the location of the ERP signal than did the Woody procedure. For data with sufficient signal-to-noise ratios, both techniques performed approximately equally. At the highest signal-to-noise

ratio examined here (Figure 5-6), there is the suggestion of an advantage for the Woody procedure.

o Figures 5-7 through 5-9 — Effect of jittering a medium-sized ERP over three different latency ranges. Both techniques performed reasonably well at all degrees of jitter. However, as the degree of jitter in the simulated ERP was increased, the Aunon/McGillem technique developed a bias in the latencies chosen. The explanation for why this occurred is not apparent at present, and further exploration is needed to determine if this effect is artifactual or not.

o Figures 5-10 through 5-12 — Effect of introducing a second ERP component at three different mean distances (in latency) from a first ERP component. As expected, the Aunon/McGillem technique fared better than the Woody technique at reproducing waveforms in which more than one ERP component occurred. At the signal-to-noise ratio used here, the Woody procedure did a good job of detecting the first (and larger) peak. The second peak was consequently diluted in the Woody latency-corrected average. The Aunon/McGillem procedure's ability to identify the second (and smaller) peak increased as that peak occurred at a further distance in time from the first peak.

o Figures 5-13 through 5-15 — Effect of varying the frequency composition of the background EEG, simulating more or less alpha band activity in the EEG. The Aunon/McGillem technique was unperturbed by the presence of relatively more activity in the background EEG in the frequency band of the ERP signal, while the Woody procedure performed somewhat less well as the spectrum of the background EEG impinged on the signal.

These results must be treated as preliminary and suggestive rather than definitive. The manipulations performed here are representative of the sort of analyses that need to be performed in more detail, and the present results provide encouragement that the approach implemented here is a fruitful one.
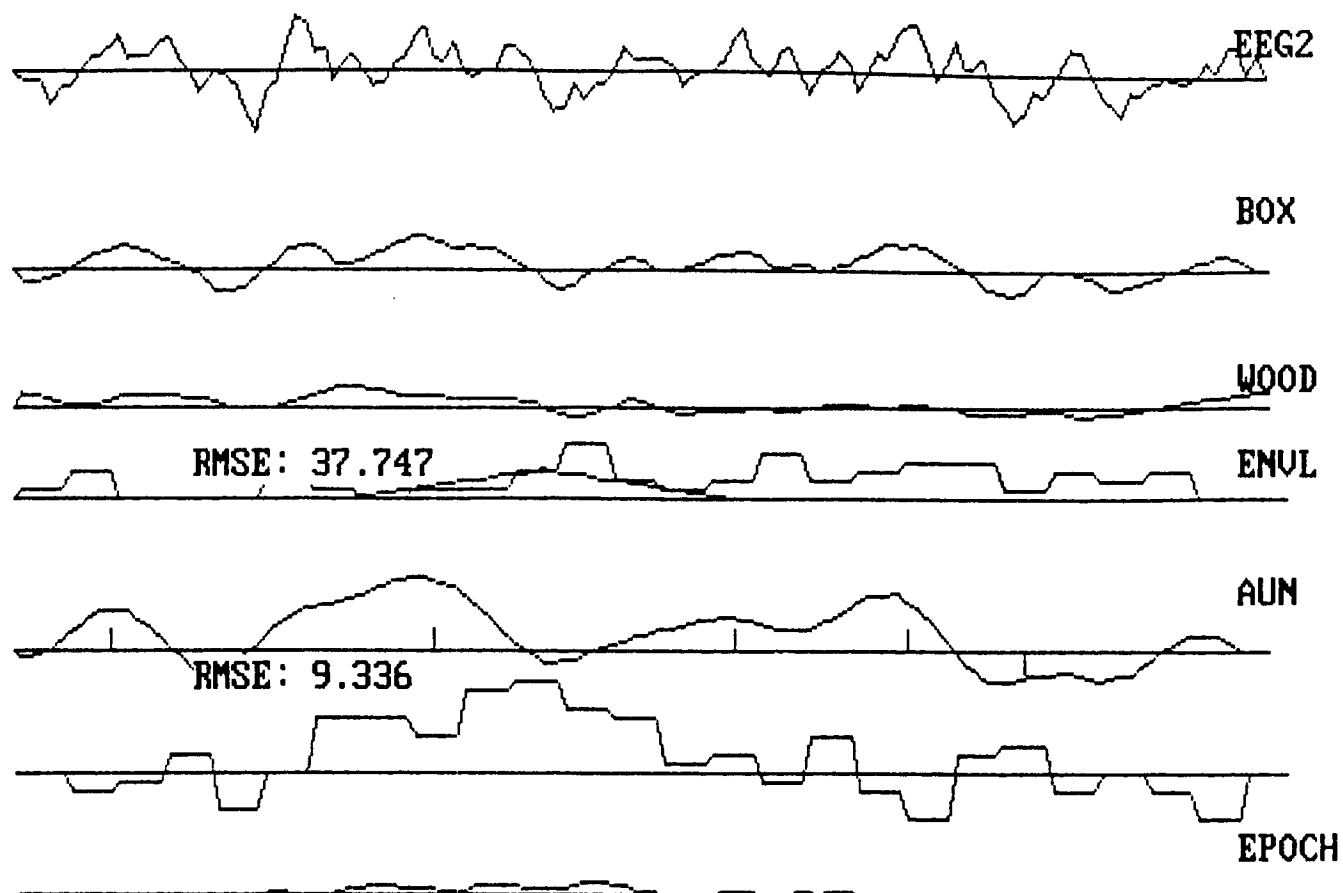
Figure 5-4. ERP amplitude manipulation, low signal-to-noise.

Conditions:

    Typical background EEG (1280 msec epoch)

    Single ERP of low amplitude

    50 trials

    Latency jitter of 400 msec

Note:

    Woody technique does poorly at finding the ERP signal; Aunon/McGillem
    technique does reasonably well.

EEG2

BOX

WOOD

RMSE: 3.697   ENVL

AUN

RMSE: 3.629
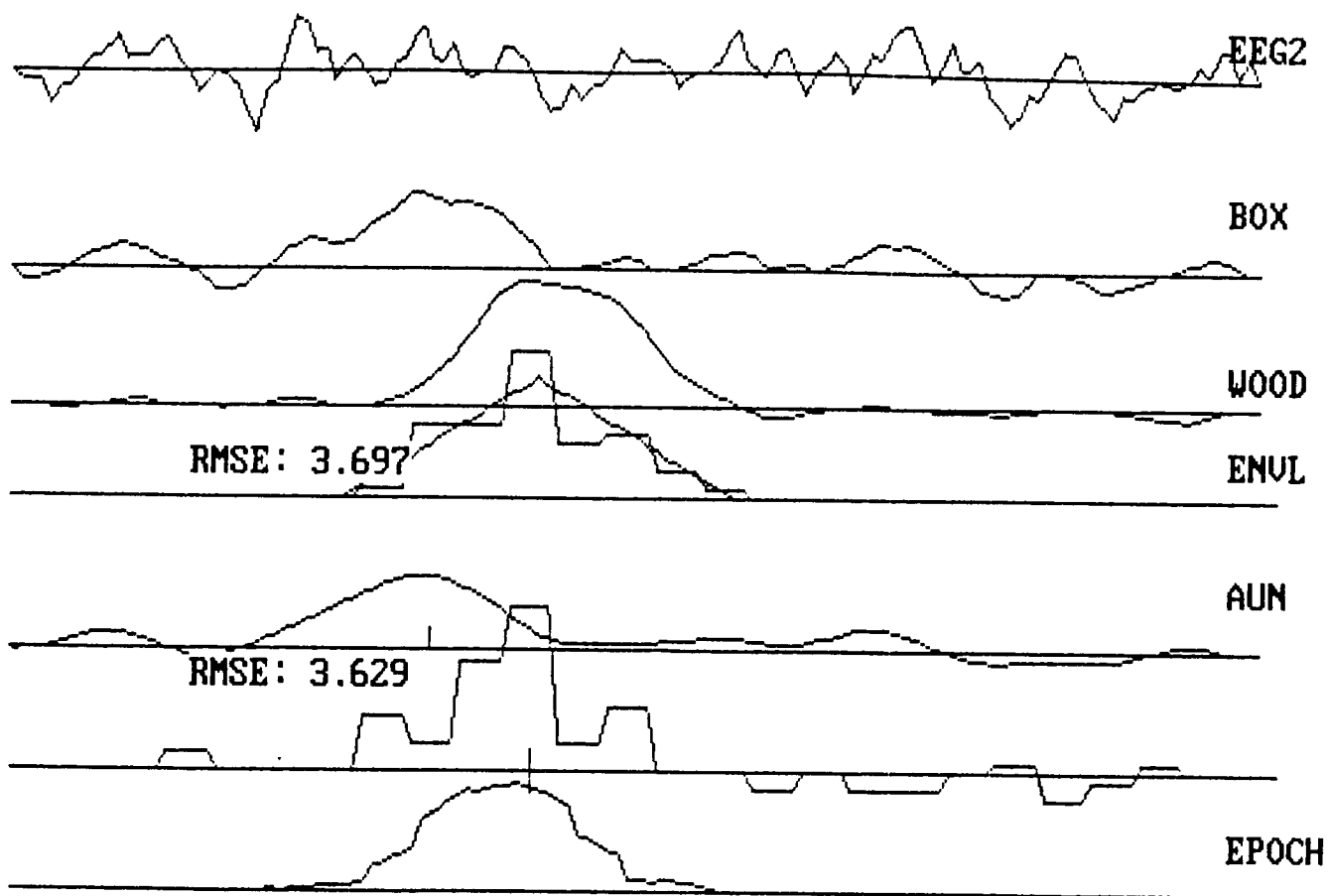
EPOCH

Figure 5-5. ERP amplitude manipulation, medium signal-to-noise.

Conditions:

    Typical background EEG (1280 msec epoch)

    Single ERP of _medium_ amplitude

    50 trials

    Latency jitter of 400 msec

Note:

    Both techniques perform well in finding the ERP signal.
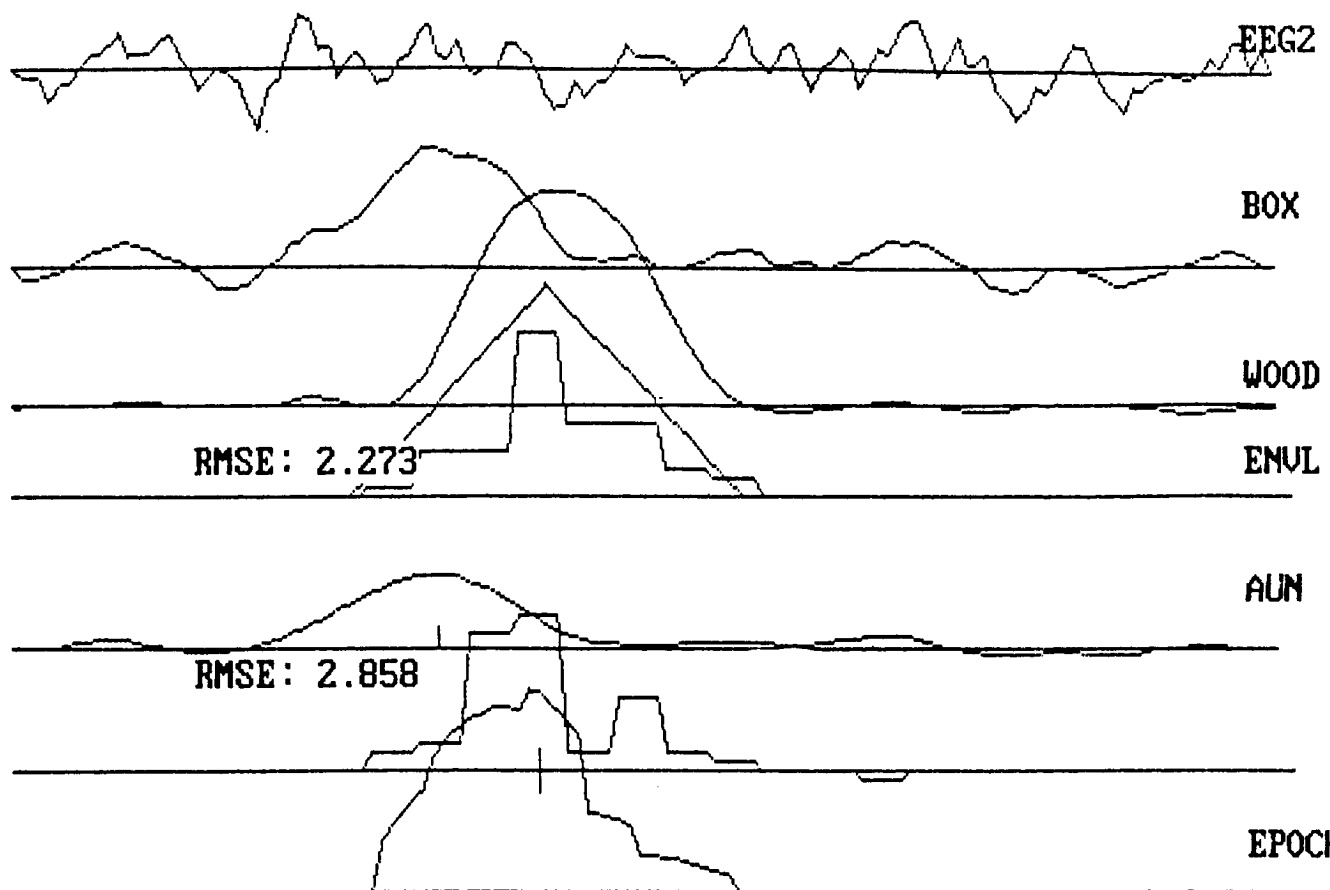
Figure 5-6. ERP amplitude manipulation, high signal-to-noise.

Conditions:

    Typical background EEG (1280 msec epoch)

    Single ERP of <u>high</u> amplitude

    50 trials

    Latency jitter of 400 msec

Note:

    Both techniques perform well in finding the ERP signal. RMSEs are still lower (indicating better performance) than with medium amplitude signals. Interestingly, there is the suggestion of an advantage for the Woody procedure.

EEG2

BOX

WOOD

RMSE: 2.677   ENVL

AUN

RMSE: 3.786
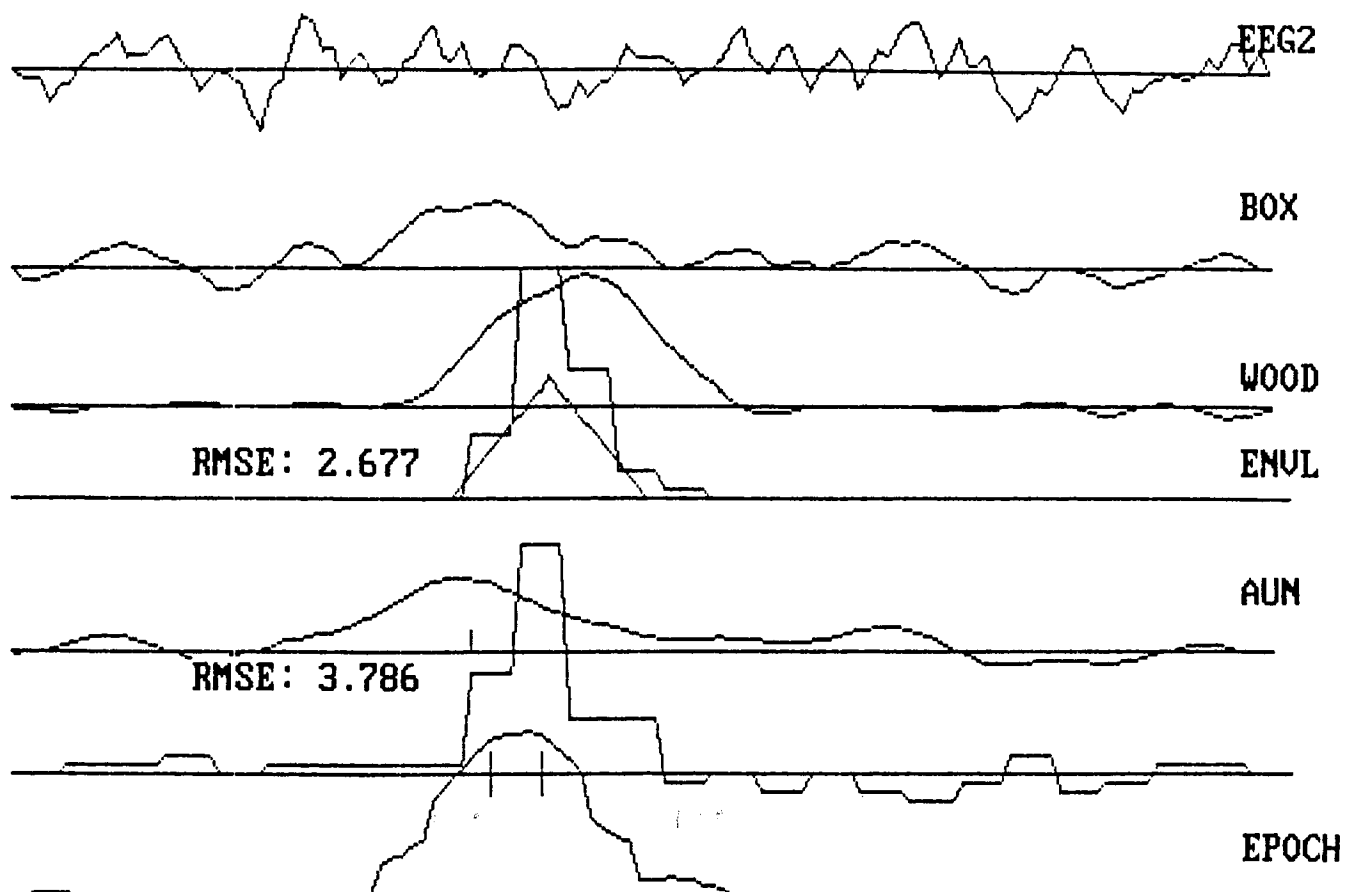
EPOCH

Figure 5 7. ERP jitter length manipulation, narrow distribution of
latencies.

Conditions:

    Typical background EEG (1280 msec epoch)

    Single ERP of medium amplitude

    50 trials

    Latency jitter of 200 msec

Note:

    Both techniques perform well in finding the ERP signal, with the suggestion
    of an advantage for the Woody procedure.

EEG2

BOX

WOOD

RMSE: 3.697   ENVL

AUN

RMSE: 3.629
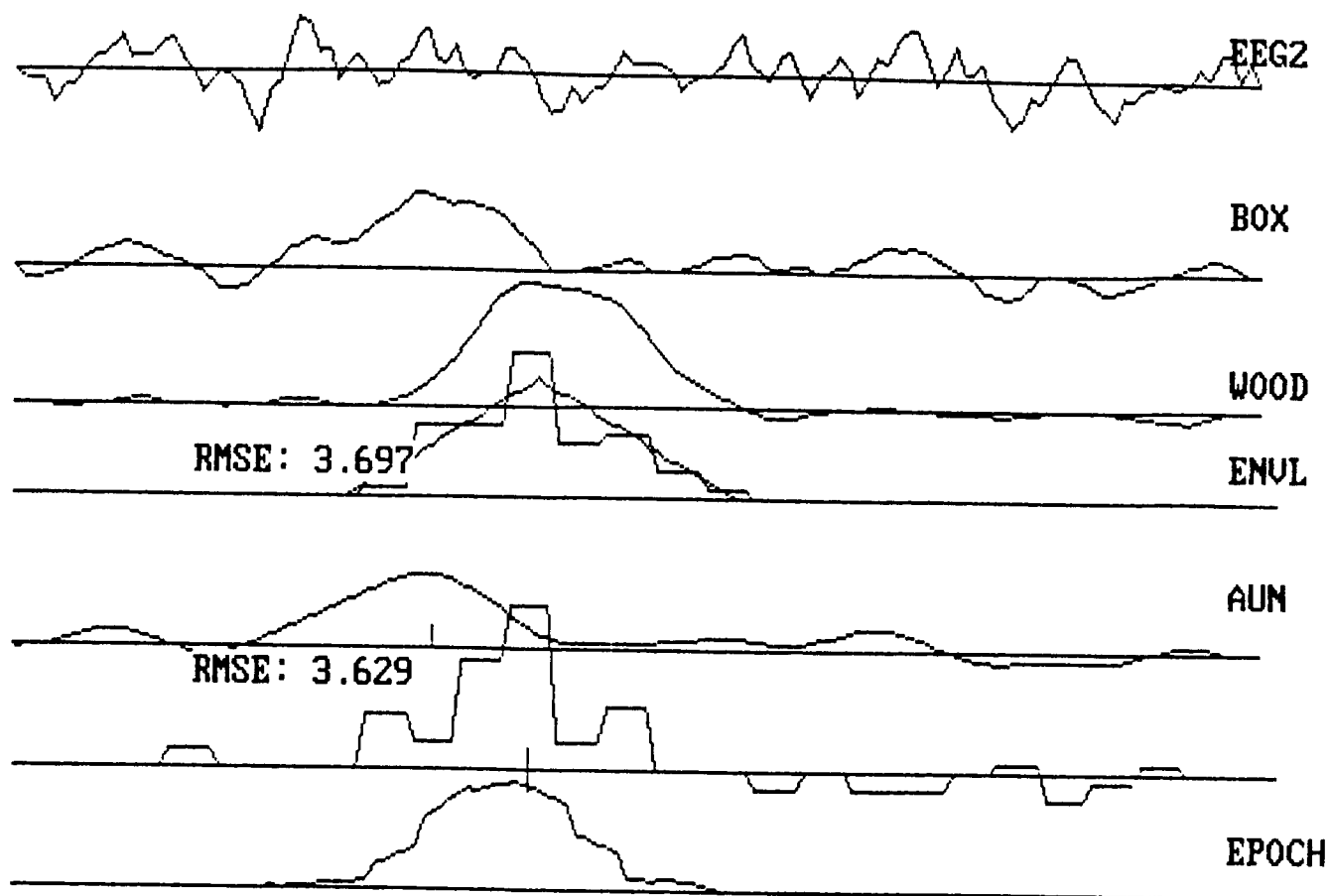
EPOCH

Figure 5-8.   ERP jitter length manipulation, medium distribution of
              latencies.

Conditions:

    Typical background EEG (1280 msec epoch)
    Single ERP of medium amplitude
    50 trials
    Latency jitter of 400 msec

Note:

    These conditions are the same as those in Figure 5-5.   Both techniques
    perform well in finding the ERP signal.

Figure 5-9. ERP jitter length manipulation, wide distribution of latencies.

Conditions:

Typical background EEG (1280 msec epoch)

Single ERP of medium amplitude

50 trials

Latency jitter of 800 msec

Note:

Both techniques perform reasonably well in finding the ERP signal, although the Aunon/McGillem procedure has developed an offset in the latency at which it reports the signal. This bias may be artifactual.
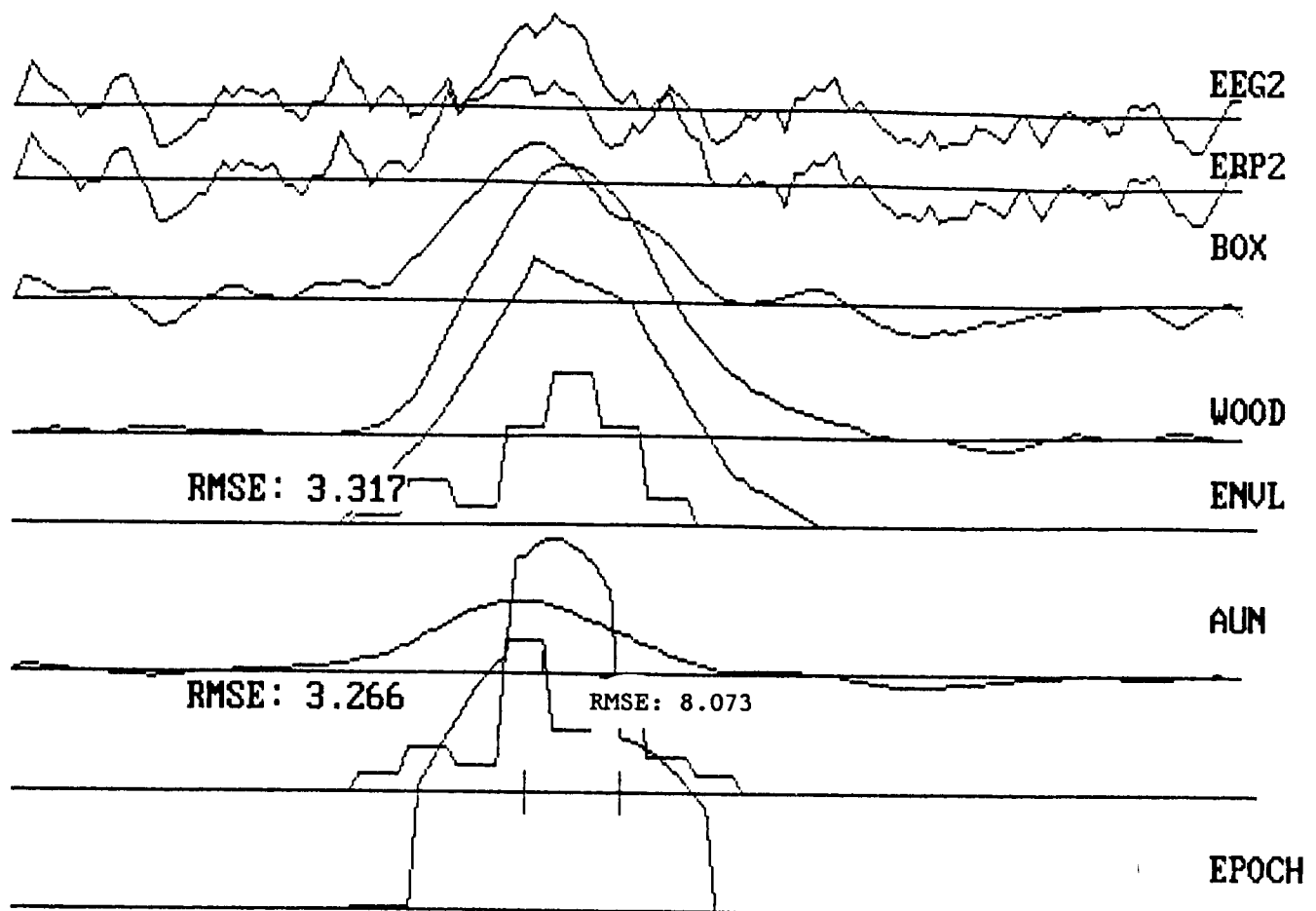
Figure 5-10. Two ERP component manipulation, small distance between the two.

Conditions:

Typical background EEG (1280 msec epoch)

Two ERP components; first of larger amplitude than the second; 100 msec mean separation

50 trials

Latency jitter of 400 msec in both ERP components

Note:

Left-most RMSE values are mean latency error with respect to the first ERP component. Because this ERP component was the larger of the two, both techniques tended to lock onto it.
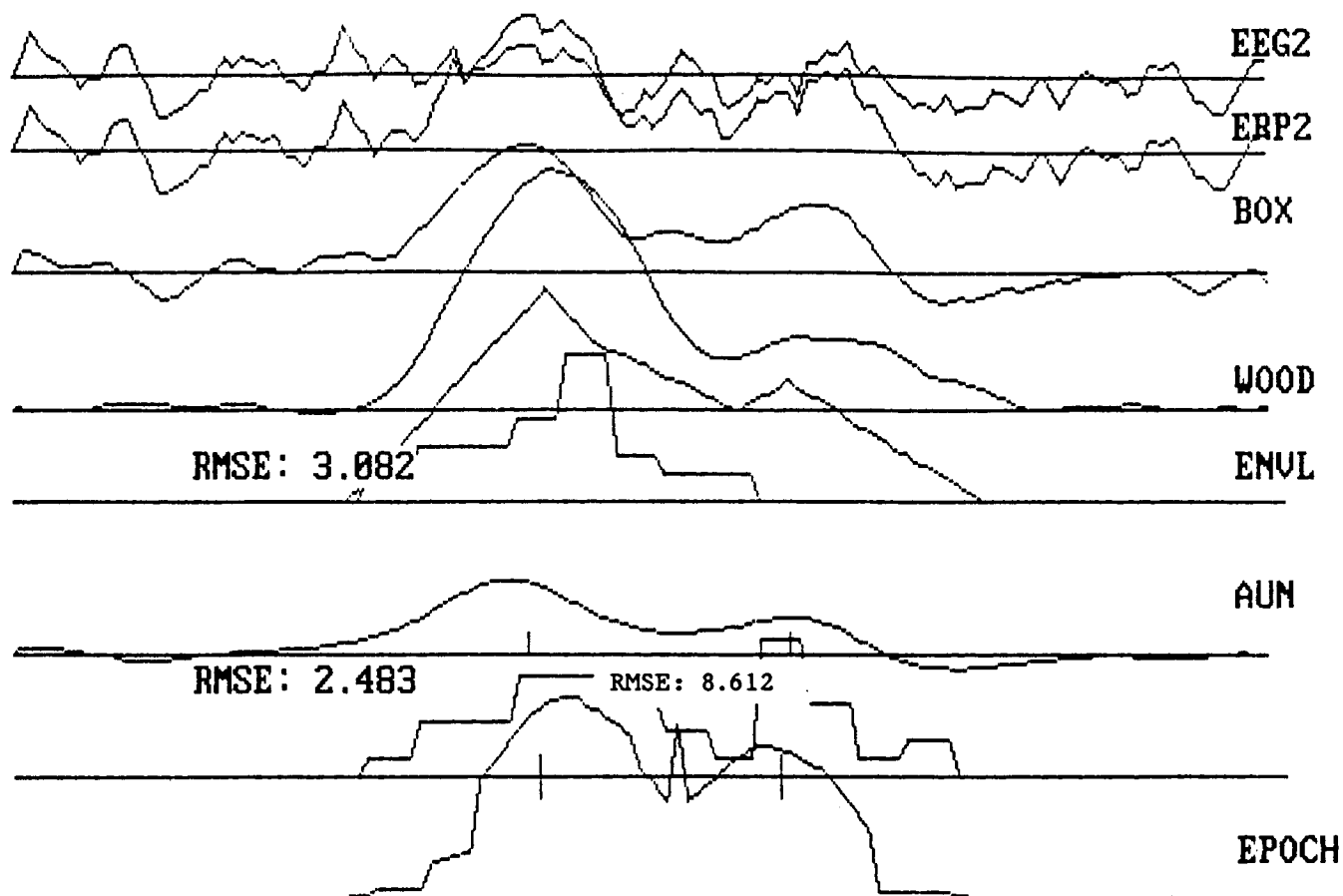
Figure 5-11. Two ERP component manipulation, medium distance between the
             two.

Conditions:

  Typical background EEG (1280 msec epoch)

  Two ERP components; first of larger amplitude than the second; 250 msec
  mean separation

  50 trials

  Latency jitter of 400 msec in both ERP components

Note:

  Left-most RMSE values are mean latency error with respect to the first ERP
  component.   Because this ERP component was the larger of the two, both
  techniques did well at detecting it. In the region of the second peak, the
  Woody latency-corrected average presents a "smeared out" rendition of the
  underlying waveshape, while the Aunon/McGillem latency-corrected average
  provides a more accurate representation of this second component.
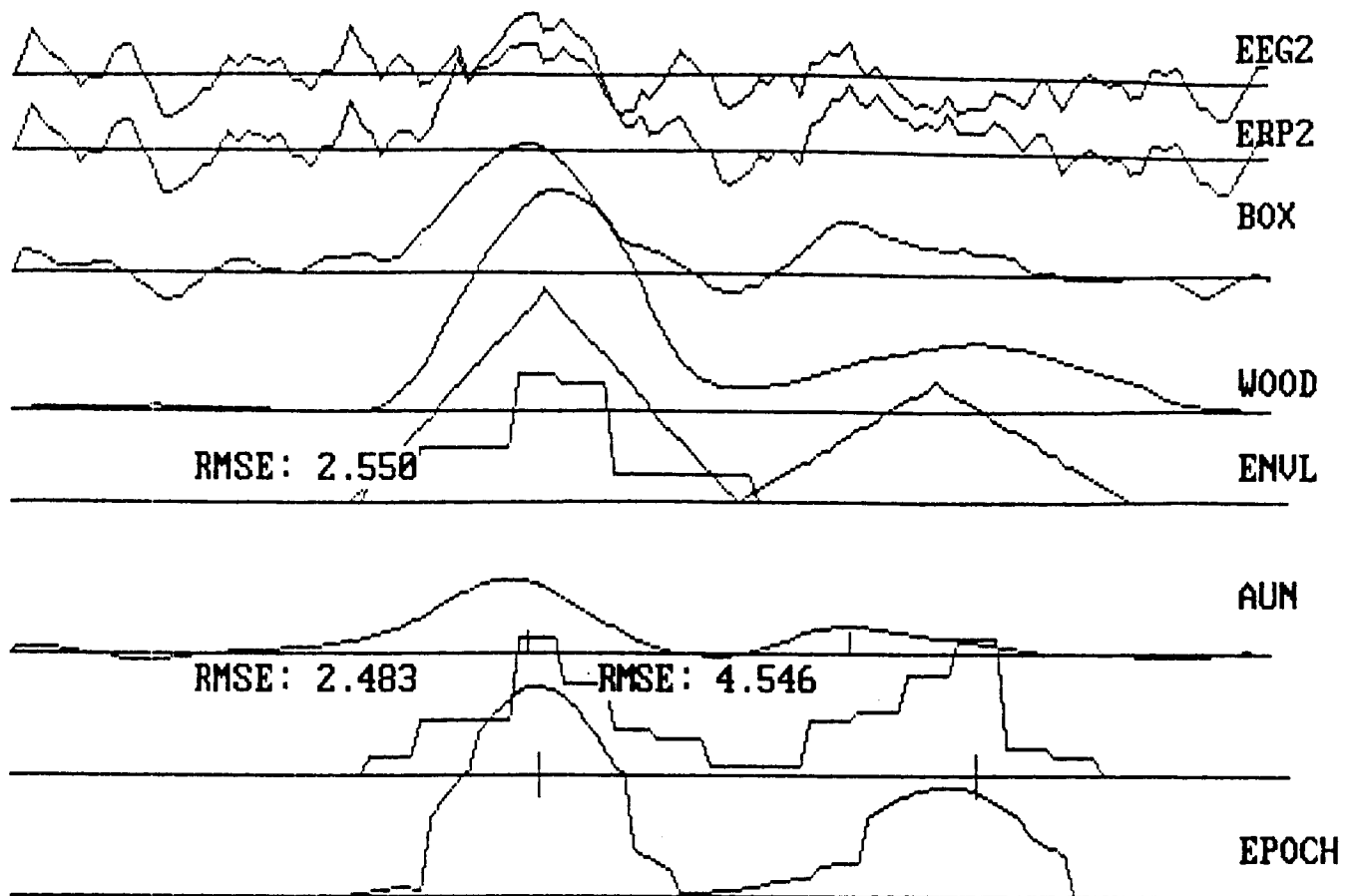
Figure 5-12. Two ERP component manipulation, large distance between the two.

Conditions:

Typical background EEG (1280 msec epoch)

Two ERP components; first of larger amplitude than the second; 400 msec mean separation

50 trials

Latency jitter of 400 msec in both ERP components

Note:

Left-most RMSE values are mean latency error with respect to the first ERP component. Because this ERP component was the larger of the two, both techniques did well at detecting it. In the region of the second peak, the Woody latency-corrected average presents a "smeared out" rendition of the underlying waveshape, while the Aunon/McGillem latency-corrected average provides an accurate representation of this second component.

EEG2

BOX

WOOD

ENVL

RMSE: 3.697

AUN

RMSE: 3.629
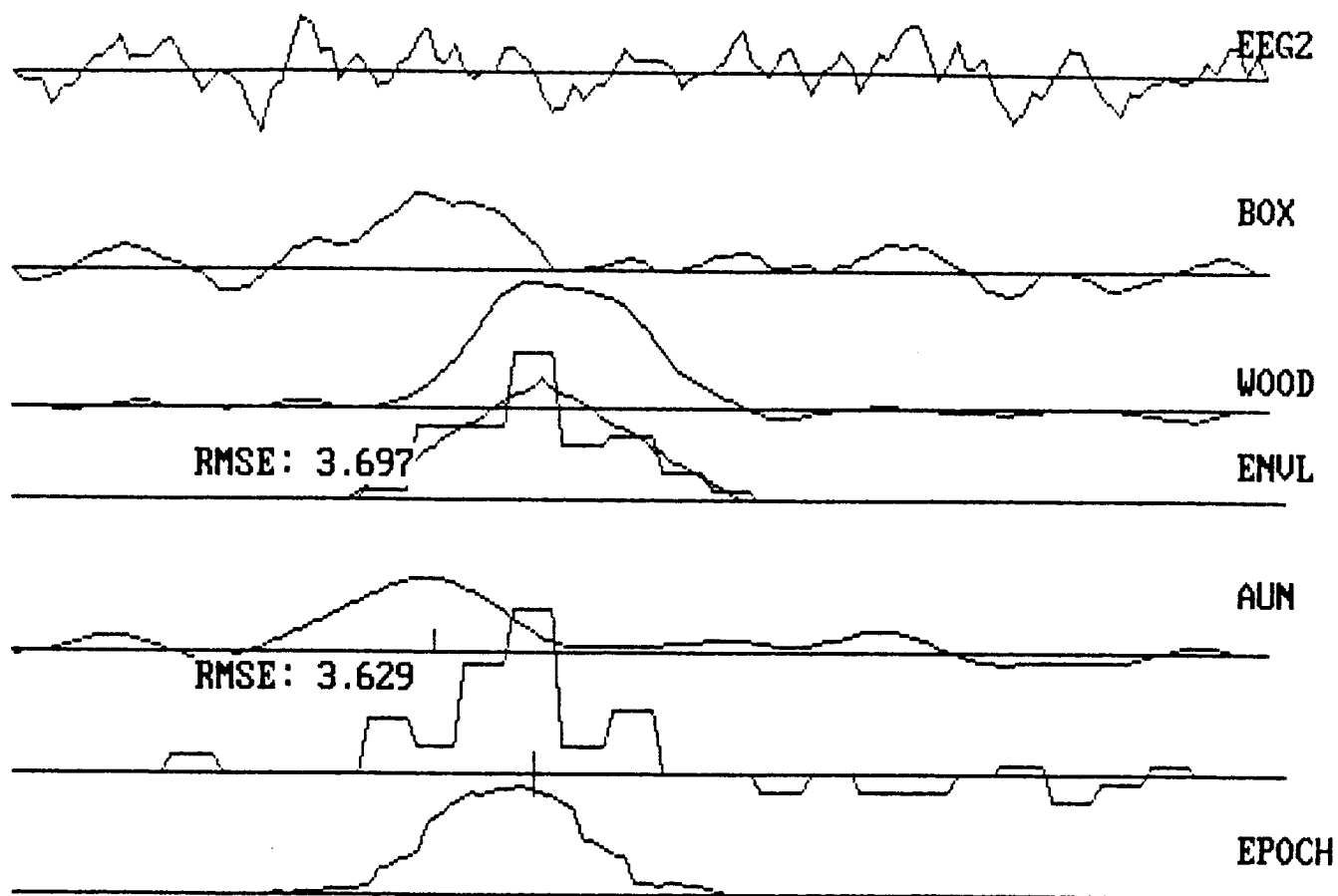
EPOCH

Figure 5-13. Background EEG manipulation, little spectral overlap with ERP.

Conditions:

Typical background EEG (1280 msec epoch); equal gains for delta and alpha frequency bands

Single ERP of medium amplitude

50 trials

Latency jitter of 400 msec

Note:

These conditions are the same as those in Figures 5-5 and 5-8. Both techniques perform well in finding the ERP signal.
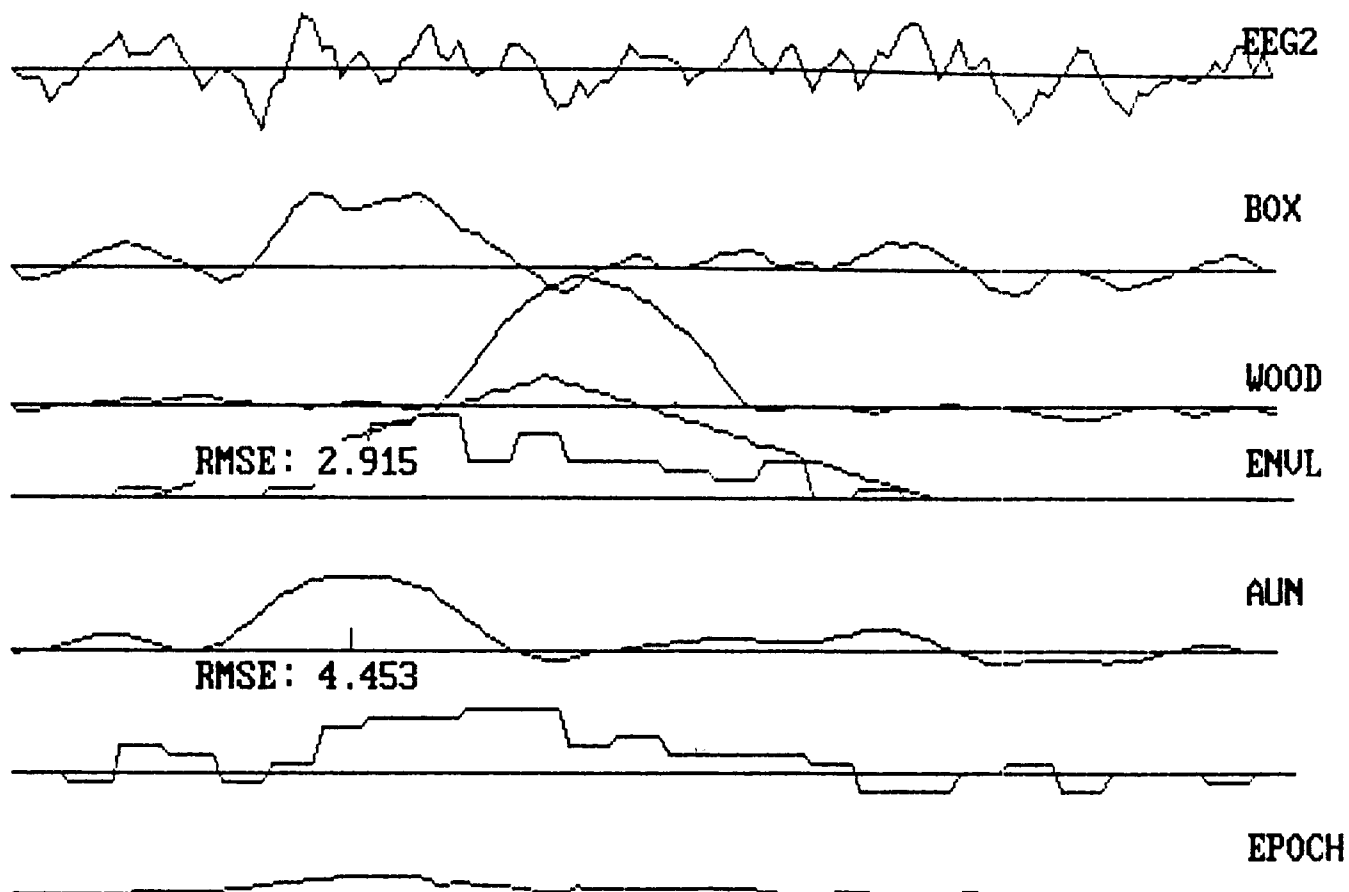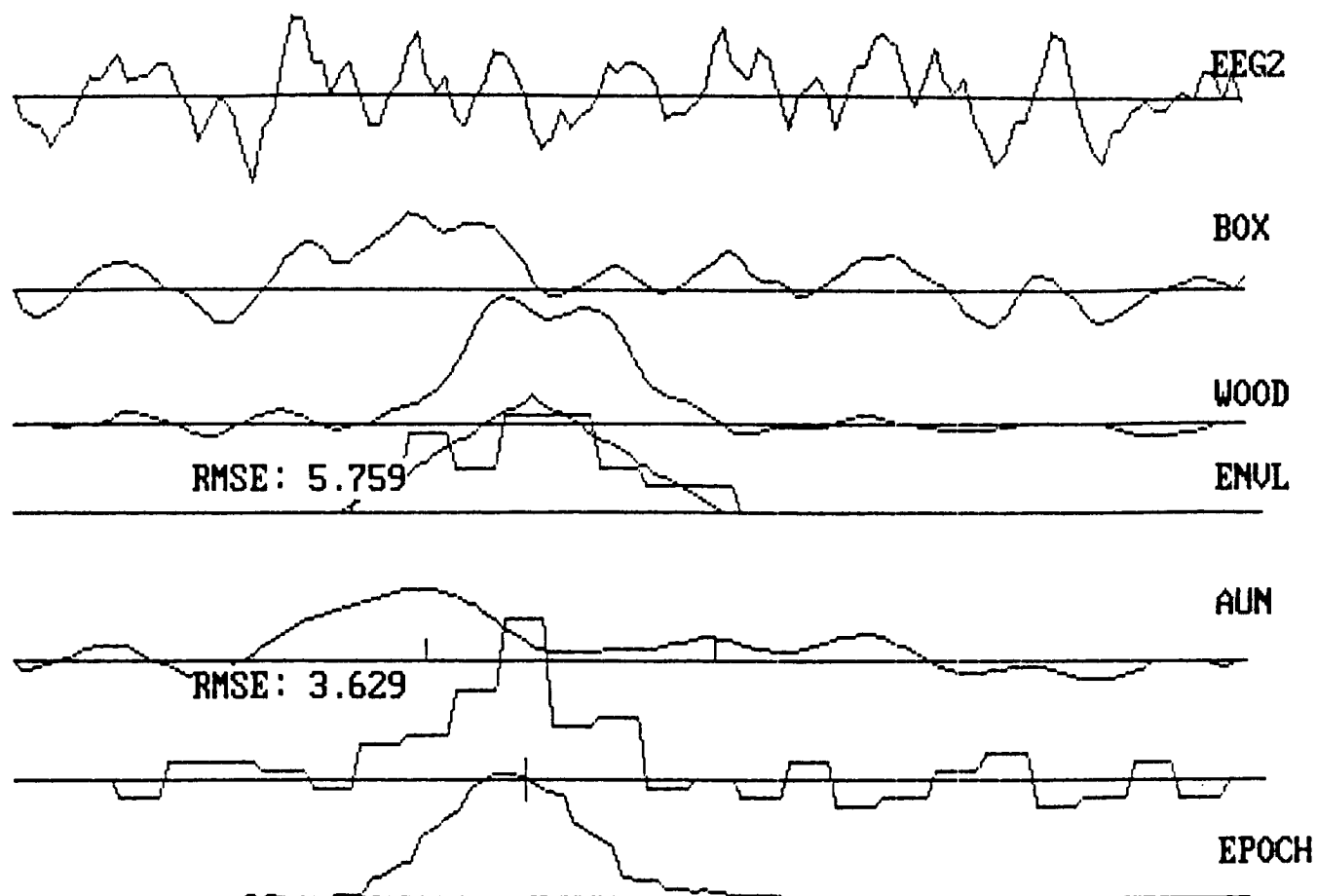
Figure 5-14. Background EEG manipulation, moderate spectral overlap with
ERP.

Conditions:

Typical background EEG (1280 msec epoch); alpha band gain twice that of
delta band gain
Single ERP with medium amplitude
50 trials
Latency jitter of 400 msec

Note:

The Aunon/McGillem procedure does as well as when there was little spectral
overlap between signal and noise (see Figure 5-13); however, the Woody
procedure does not do as well in accurately detecting the ERP signal.
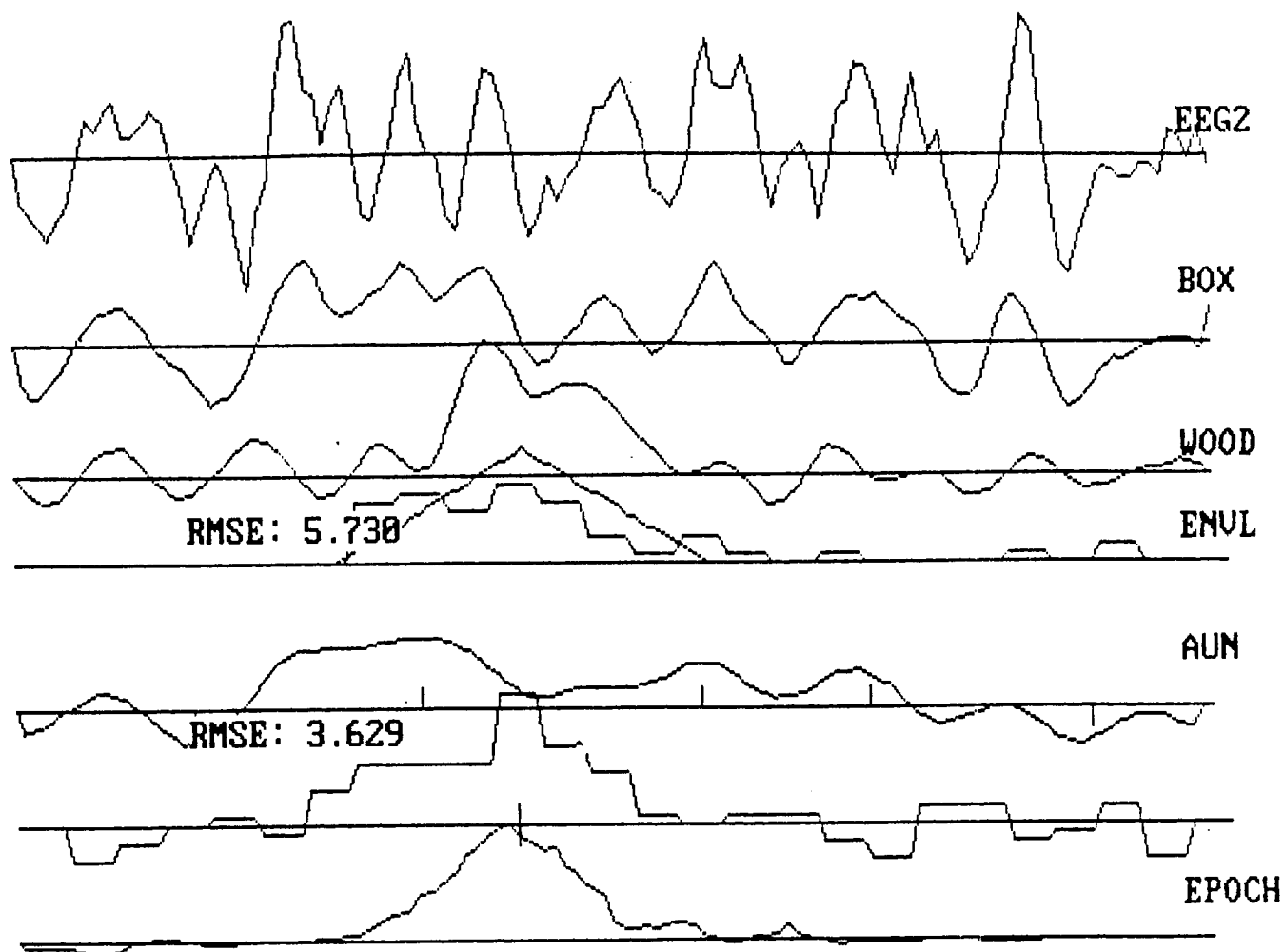
Figure 5-15. Background EEG manipulation, much spectral overlap with ERP.

Conditions:

Typical background EEG (1280 msec epoch); alpha band gain four times that of delta band gain

Single ERP with medium amplitude

50 trials

Latency jitter of 400 msec

Note:

The Aunon/McGillem procedure does as well as when there was little spectral overlap between signal and noise (see Figure 5-13); however, the Woody procedure does not do as well in accurately detecting the ERP signal.

5.5 Future Directions

It is hoped that the software tool developed here, and the approach that it embodies, will support a number of future explorations by both NASA and ARD investigators. The present implementation will be useful for exploring, in a much more extensive manner than was possible within the present scope of work, the conditions under which the Woody and Aunon/McGillem approaches are successful and the conditions under which they falter. The analyses reported in the last chapter are representative of several types of manipulations that need to be studied in more detail before firm conclusions can be drawn about the relative merits of the two techniques.

In addition, the present software implementation may serve as a starting point for future enhancements that could extend the present approach in a number of different directions. These future directions fall into several categories — enhancements to the present code that would be desirable based on our experience with the software to date, uses of segments of the present package to support other data analytical efforts, and prospects for enhancements that would qualitatively effect the algorithmic approaches or the conditions under which single-trial analysis of ERPs may be possible. In the following sections, each of these categories will be summarized in turn.

5.5.1 Useful Enhancements to the Current Version of the Software

The following modifications and enhancements to the present software would supplement the capabilities already implemented:

o   Implementation of the Aunon/McGillem MMSE filter as an option that could be applied to the simulated data prior to passing it into both analytical procedures.

o   Allow the Woody procedure to be applied in its original iterative, multi-pass form, so that the possible advantages of the adaptive feature envisioned by Woody can be more fully explored.

o   As additional statistics of merit, implement measures derived from the

latency-corrected average waveforms; these measures could include a simple amplitude measure, the ratio of latency-corrected average amplitude to actual ERP signal amplitude, and/or the cross-covariance of the latency-corrected average waveform with the ERP signal waveform.

o   Provide an option for the user to generate more than two ERP components; this would allow more realistic modeling of actual data with which an investigator may be faced.

o   Provide for the possibility of the generated ERP components being waveforms other than half sine-waves.

5.5.2   Segmentation of the Current Version for Other Data Analytic Uses

While the present implementation is optimal for the comparative analysis of simulated data, the data generation and analytical routines could also be useful in other configurations.   In particular, the output of the data generation processes may be of use for testing other ERP data analytical approaches (e.g. principal components analysis, step-wise discriminant analysis, cluster analysis) or for modeling EEG/ERP data.   Conversely, the power and flexibility of the present Woody and Aunon/McGillem routines should be useful for processing real data in a manner quite analogous to the present analysis of simulated data.   These possibilities argue for either implementing segments of the present code in new programs, or providing additional options for output of generated data and input of "live" data at appropriate intermediate stages of the present analyses.   The modularity with which the present software has been designed will serve such new applications well.

5.5.3   Prospects for Qualitative Enhancements to the Present Single-Trial Analysis Approach

As mentioned previously, one advantage of the approach taken here is that it may encourage the development of a single-trial analysis approach that works better than either the Woody or Aunon/McGillem approaches as originally implemented.   In this vein, there are several qualitatively new directions in which the present explorations may lead:

o   In that both the Woody and Aunon/McGillem techniques are applied to only
    one channel of scalp data at a time, they both ignore a valuable source of
    additional information, namely characteristic differences in ERP amplitude
    across the scalp. Recently, a technique has been suggested for
    incorporating scalp distribution information (Gratton, et al., 1989).
    Termed "Vector Analysis," this technique takes amplitudes across an array
    of scalp sites and uses them to define a line in multidimensional space, in
    which each axis corresponds to the amplitude at one scalp site. Thus, if
    the scalp distribution of a component of interest (e.g., P300) is known in
    advance, this approach can be used as a filter to quantify the extent to
    which the data at a given time point match this known distribution.
    Gratton, et al., (personal communication) have applied this technique for
    detecting P300s on a single-trial basis and found it to be superior to
    either the Woody procedure or to a simple peak detection algorithm. The
    Vector Filter, in fact, proved useful as a pre-treatment to the data prior
    to doing a conventional Woody analysis. This approach has not, however,
    been compared to the Aunon/McGillem procedure. It would be interesting to
    add a Vector Filter option to the present software and to more fully
    explore its virtues with simulated data.

o   Aside from the latency correction of conventional ERP averages, a more
    general single-trial analysis issue involves the detection of ERP
    components in ongoing EEG without prior knowledge of the timing of the
    eliciting stimulus. If the present pattern recognition algorithms, with
    appropriate pre-filtering of the data, should prove so powerful as to
    provide such a means for identifying ERPs in ongoing EEG, there would be
    enormous implications for the application of this technology in operational
    settings. One could, in effect, forego, or at least be less dependent
    upon, the "artificial intelligence" that is presently required in
    uncontrolled environments to detect the fact that a stimulus of potential
    interest has occurred. The software developed here, perhaps in conjunction
    with enhancements such as the Vector Filter (and consequent modifications
    to the data generation routines to simulate concurrent activity at
    different scalp sites), can provide the basis for exploring such issues.

o   Related to the above two sets of issues, and the present interest in

applications in operational settings, is the matter of a real-time implementation of some version of these single-trial analysis techniques. Ultimately, if a real-time ERP pattern recognition approach appears feasible, it may be necessary to implement it in special-purpose hardware or firmware, in order to accomplish the processing speed required. However, applying various algorithms to EEG/ERP data in a single pass, in non-real-time, would be the first step in exploring and developing an approach that might eventually work in real-time. The interactive philosophy and displays of accumulating data which are manifest in the software developed in the present project are quite consistent with an eventual real-time application. Therefore, the present software could serve as a developmental test-bed for the necessary real-time algorithms.

6.0 PREPARATIONS FOR SIMULATOR-BASED VALIDATION OF ERP MEASURES OF WORKLOAD

In keeping with the ultimate goal of transitioning ERP technology into settings where it can be used as a tool for systems engineering or on-line decision-making, the present project included a number of activities to prepare for a validation of ERP measures of mental workload in the NASA Advanced Concepts Flight Simulator (ACFS). The objective of such a validation would be to demonstrate workload-related effects on ERPs elicited unobtrusively by naturally occurring events in the ACFS during one or more realistic flight scenarios. The preparatory activities conducted here drew upon the results of the present "task analysis" of the ACFS displays and aircrew tasks as well as the results of the laboratory-based ERP studies. These activities, which involved key input from NASA Langley research scientists and simulator engineers, included the following:

o   A preliminary experimental design, choice of ACFS scenario and preliminary specification of ACFS performance measures to be viewed in conjunction with ERP measures.

o   Specification of a communication protocol by which information about the scenario, timing and nature of ACFS events, and pilot's performance would be passed from the VAX 8650 controlling the simulation to a laboratory computer recording ERPs.

o   Installation of a package of subroutines on the laboratory computer at NASA Langley to support ERP data acquisition functions.

6.1 Preliminary Specification of Experimental Design and ACFS Scenario

The objectives of an initial, simulator-based validation can best be met by limiting the study to a scenario involving a single phase of flight, in which mental workload can readily be manipulated and in which a number of display features will be exercised. For the purpose of specifying a set of such

manipulations and corresponding ACFS measures, we assume a landing scenario. It is envisioned that a series of landing runs would be "flown" by each subject, with workload being manipulated across runs by varying simulated wind conditions or turbulence and/or "one engine out" versus both engines functioning normally.

ERPs would be triggered by such events as the following:

o   The onset of certain discrete messages on the screen:
    --   OUTER MARKER
    --   MIDDLE MARKER
    --   INNER MARKER
    --   DECISION HEIGHT
    --   FLARE BAR
    --   FLAPS?

o   The position of the Flight Director Ball relative to the target area in which the pilot is trying to maintain it.

o   When the Air Speed Error "bug" exceeds a specified level.

o   When discrete alarms go off, to be manually reset.

These ERPs would be coded as to the eliciting event and the workload conditions under which they were elicited.   At a minimum, the following variables should be logged by the ACFS for use as converging performance measures for comparison to these recorded ERPs:

o   Deviation from Indicated Air Speed.

o   Vertical Air Speed.

o   Deviation from Vertical Velocity.

o   Track Angle Error.

o   Deviation from Glide Slope.

o   Landing footprint:

    —   X deviation from touch down point

    —   Y deviation from touch down point

    —   Sink rate (vertical loss)

    —   "Crab angle"

o   Stick Inputs:

    —   Roll

    —   Pitch

    —   Rudder

    —   Pedals

    —   Throttle

6.2   Communication Protocol Between Simulator Computer and Laboratory Data Acquisition Computer

It was assumed that the DEC MINC computer (LSI–11/23 based) in the Crew–Vehicle Interface Research Branch at NASA Langley would be used to digitize, label and store event–related potential (ERP) data time–locked to specified display elements in the ACFS. These physiological waveforms would be digitized at a rate of 100 Hz for approximately 1.5 seconds after being triggered by the onset of a simulator event, which for this study would be a specified change in the primary flight display. Therefore, it is important that this triggering occur with, at worst, 10 msec accuracy, so that the error in time–locking ERPs would never be more than one digitized time point.

The information displayed on the cockpit CRTs, as well as the fact that an event of interest is about to occur, would be computed by the ACFS model running on the VAX 8650 and then downloaded into the Ikonas display generator. Because it is advisable to time-lock to the actual presentation of the event to the pilot, and the "repainting" of the cockpit CRTs by the Ikonas occurs asynchronously with respect to the downloading of information from the VAX, it is necessary to pick up a timing pulse from the Ikonas in order to trigger the recording of ERPs by the MINC with the required accuracy of timing.

In addition to this time-critical trigger information, it is advisable to tag each digitized waveform with a code delineating which of the numerous ACFS events of interest elicited a given ERP (i.e., caused the trigger to occur). This coding will allow the waveforms to be sorted properly during off-line data analysis.

In that the information about which event occurred will reside on the VAX, it is necessary to send an event code from the VAX to the MINC in real-time. Other information about the condition of the simulation at the time the ERP is elicited (i.e., values of specified global variables indicating flight parameters, pilot control actions, etc.) can be logged on the VAX and combined with the ERP data off-line. A common time code is necessary to allow this off-line combination of condition data from the VAX and ERP data from the MINC, so the VAX should also pass the MINC, in real-time, an indication of elapsed time in addition to the event code. However, unlike the transmission of the trigger pulse from the Ikonas to the MINC, this transmission of event and elapsed time information from the VAX to the MINC need not be strictly synchronized with the recording of an ERP. At present, it appears that the transmission of condition information from the VAX to the MINC should precede the transmission of a display update from the VAX to the Ikonas, and consequently the transmission of the trigger pulse from the Ikonas to the MINC.

The oculometer which is being implemented in the ACFS is another source of input to the MINC. Most of the outputs on the oculometer are analog signals which can be digitized by the MINC along with EEG and EOG data coming from the pilot. However, there is one digital output from the oculometer, a data quality flag. Obviously, changes in this source of digital input to the MINC

6-4

will occur asynchronously with respect to both the timing information coming from the Ikonas and the condition information coming from the VAX 8650.

6.2.1 Ikonas/MINC Link

The Ikonas display generator is a raster-scanning device with two display buffers. The information in one buffer, referred to as an "animation frame," is displayed on the CRT while the information in the other buffer is updated with a transmission from the VAX. The two buffers then switch, so that the animation frame that has already been displayed can be updated with a subsequent transmission from the VAX. Each display screen, referred to as a "video frame," is painted in an interlaced manner, such that it takes 1/60 of a second to display one field (every other line) of the screen and 1/30 of a second to display an entire video frame. The number of video frames per animation frame is dependent on the speed at which animation frames can be generated. The Ikonas is presently set to display two video frames before switching buffers to a new animation frame.

It is necessary to time-lock the onset of an ERP to the actual occurrence of the eliciting event on the display screen. In any given scenario there will be numerous such events of interest and they will occur at different places on the screen. The approach proposed by NASA Langley engineers to achieve 10 msec accuracy in this time-locking is to draw an invisible cursor into a spare bit plane, at the place on the screen that the event of interest will occur (visibly on one or more of the other bit planes), and to send out a pulse on the cross-bar switch bit when the raster strobes the pixels that comprise this cursor. Cursors at different positions would be drawn for different events, and which, if any, cursor is drawn on a given animation frame would depend on the information being downloaded from the VAX. The onset of a signal on the cross-bar switch will thereby correspond, within nanoseconds, to the onset of the event of interest on the screen. In that this signal may occur only while the raster is scanning over the pixels which compose the cursors, it may be necessary to send this signal through a one-shot logic gate in order to shape it into a signal that will be appropriate for digital input to the MINC.

6-5

6.2.2 VAX/MINC Link

In order to accommodate the transmission of event number and elapsed time from the VAX to the MINC, a serial communication link was considered and rejected, largely due to the reportedly slow speed of the MINC in handling serial data inputs. Instead, a parallel transmission link can be fashioned by using "discrete variables" which the ACFS model uses for controlling indicator lights and other binary switches in the cockpit. A bank of these lines from the VAX which are currently not being used in the cockpit, or which are expendable, can be wired to digital input bits on the MINC. The discrete variables would then be used to send out TTL level signals on these lines, in a pattern which constitutes a meaningful binary code.

It should be possible, in one parallel transmission, to have this code convey both the information as to which event is about to occur and a time code from which elapsed time can be derived. It is likely that a typical experimental design will entail the triggering of ERPs by approximately a dozen different events (e.g., onset of CRT messages such as "OUTER MARKER"; drift of the Flight Director Ball out of the target area, drift of the Air Speed Indicator out of a pre-defined region of its scale). NASA Langley engineers have suggested that these events be entered into a look-up table on the VAX, such that when the simulation model calculates that an event of interest is to occur on the next display update, a numeric code associated with that event will be looked up and output to the MINC.

The VAX keeps time in terms of a frame counter, which increments with each updating of the simulation model (presently, approximately 16 times per second). The easiest way to pass elapsed time from the VAX to the MINC will be to alternately set and reset a single "discrete," as each new frame update occurs The MINC can then keep an incremental count of these frames, as does the VAX. By thus coding time, it will be possible to pass information about the triggering event and the corresponding time of its occurrence in a single parallel transmission from the VAX to the MINC. If eight bits of information (i.e., eight discretes) are available on the VAX, seven can be used for coding event information (thus providing for up to 127 different events, should anyone ever need that many) and the eighth can be used for transmitting the timing

code.

Other ACFS condition information will also be useful for sorting and interpreting the ERPs, but there is no need to pass this additional information to the MINC in real-time. Some of this additional condition information will remain the same for an entire scenario (e.g., workload level, wind condition, turbulence condition), while other condition variables will reflect the moment-to-moment performance of the man-machine system (e.g., air speed, altitude, stick position). This condition information will reside in global variables in the VAX simulation software and will be logged into a data file on the VAX throughout a scenario, written to mag tape at the end of a scenario or experimental session, and combined with the ERP data off-line. A time code will be stored along with this logged data and will provide the means for linking the ERP data from the MINC with corresponding condition information from the VAX.

The scheduling of the output from the VAX to the MINC could be timed to occur, in the duty cycle of the simulation, either before or after the transmission of the corresponding screen update from the VAX to the Ikonas. Because of the asynchrony between the VAX sending this information and the Ikonas displaying the new frame, it appears preferable to have the VAX send the event code to the MINC after it has calculated the screen update on which the event of interest will occur, but before it transmits this information to the Ikonas. In addition to the screen update information that is presently passed from the VAX to the Ikonas, it will now be necessary to also pass a code that designates the event which is about to be displayed, and consequently which cursor the Ikonas should draw in the background.

6.2.3 Oculometer/MINC Link

The outputs from the oculometer that is presently in use at NASA/Langley are as follows:

o   X and Y coordinates - two analog channels, indicating the location of gaze with respect to a predefined origin.

o   Pupil diameter — one analog channel, indicating the degree of pupil dilation when positive and a variety of error conditions when negative.

o   Gaze window — one analog channel, with incremental values separated by 0.3 volts over a range of 10 volts, indicating in which of approximately 30 predefined windows the user's gaze falls.

o   Data quality flag — one digital channel, indicating whether or not the eye is in track.

The oculometer thus provides four analog channels, in addition to the EEG and EOG channels, that can be input to the A-to-D converter on the MINC and sampled at 100 Hz.

6.2.4 Configuration of MINC Digital Input Bits

Based on the design implied by the above information, we presently envision the configuration of the MINC parallel (digital) input register to be as follows:

Bits 0-6   — ACFS discretes carrying condition information from the VAX
Bits 7   — ACFS discrete carrying time code counter from the VAX
Bits 8-13   — Unused
Bit 14   — Data quality bit from the oculometer
Bit 15   — ERP trigger pulse from the Ikonas

Changes in the value of this digital input register would send an interrupt to the program running on the MINC. The LABPACK software that ARD has installed on the MINC contains subroutines to service these interrupts and an application program can be written to read and interpret this register, as well as clear it appropriately to enable further input.

This program running on the MINC would make the following assumptions regarding digital input:

o   That an event code from the VAX will always precede a trigger pulse from the Ikonas.

6-8

o   That the subsequent trigger pulse will never occur more than 128 msec (two animation frames) after the arrival of the event code.

o   That more than one event can occur in a given animation frame.

o   That a different event can occur on the animation frame immediately following a preceding event, and that any event which occurs during the approximately 1.5 second recording epoch of an ERP should be noted.

6.2.5   Some Remaining Questions for Consideration

The primary unresolved issues are the following:

o   The range of times that can be expected from the arrival of the parallel input from the VAX until the arrival of the trigger pulse from the Ikonas; it appears that the minimum time would be the duration of the transmission of a display update from the VAX to the MINC; the maximum time would be this time plus one animation frame time plus one video frame time.

o   The nature of the crossbar switch output and how it should be shaped to avoid spurious trigger signals.

o   Whether the proposed communication interface can accommodate and distinguish multiple events in a given frame and different events on adjacent frames, or whether there will be some constraints on the temporal resolution for detecting and coding events.

6.3   LABPAK Software Routines for the MINC Computer

The use of general-purpose microcomputers to control psychophysiological data collection is now widespread, and hardware from numerous vendors has been used in accomplishing this goal.   Whereas hardware costs have continued downward, as new technology has become available and competition in the marketplace has increased, software development costs have increased because of the labor-intensive nature of this process.   Given the status of psychophysiological research in general, and that of event-related potentials (ERPs) in

particular, it does not seem advisable to saddle researchers with turn-key systems that limit their creativity and flexibility to implement new data collection protocols or data analytic procedures. However, productivity is greatly increased by the availability of a package of general purpose subroutines that perform the basic functions which are characteristic of virtually every psychophysiological research study. Such functions include starting and stopping timers, activating analog-to-digital (A-to-D) conversion, and issuing and receiving digital input and output (I/O), respectively. Routines that perform these functions can be implemented in a "canned" manner and then integrated, by the sophisticated programmer, into main programs that address specific experimental goals. The availability of such a package of "low-level" functional subroutines allows the programmer of such main programs to work in a higher level language, without worrying about the need to program the device handlers for each experiment.

The LABPAK subroutines which ARD installed on the MINC computer at NASA Langley were developed to provide this type of flexible capability for data collection and experimental control. They are written in MACRO-11 assembly language for use on Digital Equipment Corporation processors. The present version of these routines was adapted to run specifically on the MINC (LSI 11/23-based processor) computer system in the Crew Vehicle Interface Research Branch at NASA Langley. Typically these routines would be called from FORTRAN main programs of the user's own design, with the LABPAK library being linked with the FORTRAN source code and system specific FORTRAN library after compilation.

LABPAK provides the FORTRAN programmer with control over various peripherals (i.e. the real-time clock, the A-to-D converter, and the digital I/O interface). Moreover, it allows for the functions performed by these devices to be executed in the "background," in a somewhat asynchronous manner from the programmer's point of view. LABPAK is based on interrupt service routines which execute whenever either the real-time clock or digital interface produces an interrupt. Most of the LABPAK routines are used by the FORTRAN programmer to set parameters which direct how the interrupt service routines process the interrupts produced by these devices. During the time when the interrupt service routines are not executing, the FORTRAN program has control of the processor and may execute other tasks such as monitoring the status of various

flags set by the interrupt service routines, accessing disk files, manipulating other peripherals, or performing numerical calculations. In this way, data collection, timing, and digital I/O functions may take place in "parallel" with other program functions.

Specifically, LABPAK contains assembly language subroutines for performing the following functions:

o   Programmable Clock Handler:

--   Start the programmable clock.

--   Start software timers.

--   Turn off a software timer.

--   Turn off the programmable clock.

--   Clock interrupt service routine.

o   Analog to Digital Converter Handler:

--   Perform an A-to-D conversion.

--   Stop a direct memory access A-to-D conversion.

--   Start the digitizer cycling through a pre-stimulus wrap-around buffer.

--   Stop the digitizer and unwrap the buffer.

--   Write a marker to an active A-to-D buffer.

--   A-to-D clock interrupt service routine.

o   Digital Input/Output Handler:

-- Enable digital inputs (all bits).

-- Enable digital inputs for given bits.

-- Wait for an external digital input.

-- Disable digital input reporting on given bits.

-- Turn off digital inputs.

-- Set a specific digital output bit.

-- Turn off a specified digital output bit.

-- Write a value to the digital output register.

-- Clear the digital output register.

7.0 DESIGN SPECIFICATIONS FOR A GENERAL-PURPOSE MENTAL WORKLOAD ASSESSMENT SYSTEM

The culmination of the present Phase II SBIR project was a design specification for a computerized workload assessment subsystem that could be integrated into a variety of man-machine settings of interest to NASA. The immediate context for the design of such a system is the NASA Advanced Concepts Flight Simulator (ACFS) (see Sexton, 1983). Although the ACFS is a state-of-the-art transport cockpit simulator, the approach which underlies it is representative of the simulators available for many other man-machine systems in which the measurement of human performance is of interest. Moreover, as discussed below, the ACFS does not, as yet, have well-established means for dealing with the performance data it logs, let alone any systematic analyses for mental workload. The system specified here would provide this capability. In addition, it would provide a generic workload assessment system with the tools available to accommodate other man-machine environments of interest to NASA, both simulators and selected operational systems.

7.1 Overview of the Proposed System

As detailed herein, the hardware components and software tools now exist to warrant the cost-effective development of a prototype, microcomputer-based workload assessment system that will provide for state-of-the-art measurement of human performance in naturalistic task environments. The system that we propose will involve a powerful, personal computer work-station which could be transported and rapidly set up in locations of interest. Such a system will provide the means to amplify and record event-related potentials (ERPs) and other psychophysiological indices of workload, as well as subjective and behavioral data, from a system operator. It will implement state-of-the-art algorithms for processing these complementary measures and for making inferences about human functionality in operational scenarios. In addition, it will provide for digital communications with a host computer controlling a simulator or operational system, so as to encode system performance indices from the platform the operator is controlling. The software will provide for

the logging and database management of these measures in real-time during operational scenarios, the calculation of derived measures of human performance, and the off-line application of data analysis tools, including statistics, to the stored data. This off-line analysis capability will include optional knowledge-based guidance for the data analyst. The user will have the ability to interact with this system in various ways, both during real-time data acquisition and off-line data analysis, and to view the encoded information with interactive color graphics.

The system will be powerful, flexible, user-friendly and readily expandable, so as to accommodate new measures of workload, or new derived variables based on existing measures, as they are proposed. Most importantly, it will work in the engineering environment, processing data recorded, for example, during realistic flight scenarios in a simulator. While this workload assessment capability is envisioned as a tool-kit for use by design engineers, it has been configured with the possibility of later enhancement for real-time measurement of workload. Such real-time capabilities would allow a modified version of the hardware to be integrated into operational systems, so as to provide on-line measures of human performance that could be used as inputs to automated task allocation or decision-aiding algorithms.

The present design specification process identified not only design requirements that the workload assessment system must satisfy, but also alternative hardware and software components that will meet these requirements as well as the ways in which the specified system will foster further engineering or research and development in the workload assessment arena.

The following sources of information were used in this design specification:

o   Knowledge of the workload assessment literature and ongoing efforts in the field.

o   Knowledge of the nature of ERPs and other psychophysiological measures, and the characteristics of recording systems that are in now in use.

o   Results of the present empirical ERP research.

o The results of the single-trial data analysis of simulated EEG data.

o The results of the task analysis to identify cockpit events to which psychophysiological recordings could be time-locked.

o Design documentation on advanced cockpits.

o Available state-of-the-art technology in electronic circuit design and miniaturization.

7.2 Background and Justification for the Approach Advocated

7.2.1 Need for a Tool-kit Approach

The conclusion that we and a number of other groups have reached in recent years is that THERE IS NO SINGLE MEASURE OF WORKLOAD THAT WILL WORK IN ALL SITUATIONS, NOR IS ONE LIKELY TO EMERGE. We thus rejected the idea of advocating a system that specifically provides just for the measurement of ERP indices of workload. On the other hand, THERE IS A DIVERSITY OF EXISTING MEASURES OF WORKLOAD (including ERPs) THAT HAVE PROVEN USEFUL FOR VARIOUS SYSTEM ENGINEERING PURPOSES, BOTH IN THE LABORATORY AND IN THE FIELD. In fact, the technology required to provide for ERP and other psychophysiological measures of mental workload will readily support a variety of these other measures.

Taken together, these conclusions suggest the need for a multidimensional, tool-kit approach, whereby a variety of measures of workload could be readily applied, in various combinations, at the discretion of the engineer or behavioral scientist user. The user will be able to choose, for particular applications, which variables to record and which workload measures to derive. This approach is not unlike that used by physicians in diagnosing an illness. Numerous clinical tests may be conducted, none of which is completely reliable in and of itself; but based on the preponderance of evidence, taken together, conclusions are drawn and interventions (in this case, design decisions) are initiated.

We therefore chose to specify the development of a man-machine performance assessment system that will support subjective, behavioral and psychophysiological measures of workload, including ERPs, and which will readily interface with a variety of simulated or fielded systems. The system will allow performance data, e.g. flight variables such as those displayed in the ACFS cockpit, to be related to variables which indicate the mental status of the crewmen.

7.2.2 Lack of Such a System at Present

No such system is currently in existence. The PEARL system that was developed at the University of Illinois under contracts from the Air Force and Environmental Protection Agency (Heffley, et al., 1985), and its successor, the Air Force's Neuropsychological Workload Test Battery (NWTB) developed by Systems Research Labs, Inc., perhaps come closest to meeting this need. The design of both these systems was driven by the desire to present contrived secondary tasks in order to elicit behavioral and physiological indices of workload. They are configured primarily for acquiring event-related data, in laboratory or simulator environments. Unfortunately, both these systems are based on somewhat outdated technology which is no longer cost effective, and and their computer architecture places severe limitations on their ability to effectively acquire, reduce and store data collected in an ongoing manner over long periods of time (minutes or hours). Furthermore, neither system is readily usable by the non-programmer engineer or researcher. Their present software does not allow for data being collected on different channels at different rates. The Air Force is currently involved in the design of an upgrade to the NWTB, but it is not yet clear what features it will encompass.

A software package known as Performance and Workload Analysis (PAWAN), which was developed several years ago by the Computer Technology Associates group at Edwards Air Force Base, is also relevant. This package, which apparently is no longer being actively used at Edwards, was developed to intercorrelate flight parameters, measures of pilot behavior, and heart rate, recorded in-flight during fighter aircraft test and evaluation. PAWAN was implemented on mainframe computers which did not facilitate interactive data analysis. From the limited information available to us, it appears that PAWAN incorporated

some of the features that we envision for the recommended workload assessment system. However, the proposed system will be more generic, interactive, and comprehensive with respect to current workload measurement techniques.

Subjective ratings techniques for workload assessment have received considerable attention in recent years. Two particularly well-developed techniques, the Subjective Workload Assessment Technique (SWAT) (e.g. Eggemeier, et al., 1982; Reid, et al., 1981; Reid, 1985) and the NASA Task Load Index (e.g. Hart, et al., 1982; Hart & Staveland, 1986), are now available for IBM-PC compatible systems. The workload assessment system specified here would accommodate these techniques in their present implementation.

7.2.3   Current Means of Handling ACFS Performance Data

There are three implementations of the ACFS -- one at NASA Langley, one at NASA Ames and one at Lockheed-Georgia. There are some differences in the implementation among these three sites. For example, the ACFS at Langley is the only one driven by a VAX 8600 computer. However, at all sites the ACFS software provides extremely powerful and flexible means for logging numerous variables from the aircraft model during a flight scenario. The user can specify which variables are to be logged for a given scenario and select, for each one, how frequently it will be sampled and stored. These data are stored into memory along with time stamps (number of elapsed update cycles thus far in the scenario) on the host system and then dumped to a hard disk. The information that can thus be saved includes time-coded flight status variables that are presented on cockpit CRT displays, the binary status of other cockpit indicators such as Caution and Warning lights, current positions of the side stick controllers, switch positions of other controls actuated by the aircrew, and condition variables such as the wind or turbulence values that are being used by the model at the moment. Therefore, the means are provided for encoding practically all of the variables that one would need to characterize platform performance, and many of the variables from which operator behavior can be derived.

However, there are not, as yet, well-established means for processing these logged variables and deriving the needed performance data. At NASA Ames, where

the first ACFS came on-line, researchers have written the logged data onto magnetic tape media, loaded it onto a mainframe and analyzed it with the INGRES data base management package and the S data analysis package. While extremely powerful, these packages are also very expensive and somewhat cumbersome to use. The workload assessment system that we specify here would provide many of these same data analytic features, in a readily usable manner, along with additional measures of operator physiology and behavior which are not feasible to implement on the same system controlling the simulation. Moreover, it offers sufficient processing power for dealing with these type data, at a fraction of the start-up costs, and with the considerable savings in time and convenience that is provided by implementation on a personal computer workstation.

7.2.4 The Advantages of Implementing a PC-Based Workstation

From the present analysis of the needs of NASA and other potential users of the proposed workload assessment system it was concluded that the proposed workload assessment tool-kit system will best be implemented on a personal computer (PC) workstation. The stunning advances in speed (cycle times now up to 25 megahertz) and miniaturization that is now available in PCs, coupled with the vast amount of relatively inexpensive software that has been written for that environment, allows one to attain the kind of processing power that until recently was associated only with minicomputers, if not mainframes. Moreover, our experience indicates that it is possible to configure a PC-based system that will meet the present needs, for a fraction of the cost (we estimate 1/10 to 1/3) of minicomputer processors, peripherals, and software. Furthermore, the architecture of chips such as the Intel 80286 and 80386 (or for that matter the Motorola 68020 and its derivatives) allow for the addressing of much more on-board memory than is possible with conventional minicomputer architectures. This random access memory allows for larger programs to be linked without cumbersome overlay structures, as well as greater buffering of data being continuously acquired. Finally, there are software tools that will be useful in the present effort that are currently available only for PCs.

For recording in simulators, where space is not constrained and it is feasible to tether the subject to the recording equipment, the use of off-the-shelf PC

hardware is quite adequate. However, in considering the need to eventually develop a subsystem that could be integrated into operational settings such as flying aircraft, the necessary design specifications become much more severe. Space will be at a premium, it may not be desirable to tether the pilot, and the effects of the avionics and environmental conditions on the recording equipment are difficult to predict. Therefore, for the near-term development of this technology, it seems reasonable to limit oneself to a system that will interface with simulators or ground-based systems and which does not entail special-purpose hardware. By providing a tool-kit for managing and analyzing human performance data that have been recorded by conventional means, the specified workload assessment system offers NASA the most cost-effective means of enhancing its workload evaluation capabilities.

The use of general purpose A-to-D converters and digital input/output registers (both serial and parallel) in the present system are envisioned in order to provide a generic capability to interface with a wide variety of potential systems. The software would include easily usable, menu-driven means for the user to configure these I/O lines for communicating with a given host system. Such features would accommodate the sort of communication protocols that we developed in the present project for the VAX 8600 and MINC data acquisition computer at NASA Langley (see Section 6). If necessary, it would be possible to provide for special-purpose hardware with direct memory access to the solid state or optical memories of future flight data recorders, or the ability to encode data in an "intelligent" way directly off the data bus of operational systems. However, by avoiding these levels of specificity and designing a generic tool-kit system, one minimizes the possibility of developing special-purpose instrumentation for inflight recordings that may be obsolete by the time it would actually be used in this ambitious manner.

7.3 Workload Measures Supported

The specified workload assessment workstation will allow the derivation of at least the following measures of operator workload, based on simulator data encoded along a time base with millisecond accuracy:

o   Subjective Workload Assessment Technique (SWAT) -- this subjective rating

technique has been used extensively over the last several years by a number of groups. PC-based software to implement the conjoint analyses involved in this approach are provided upon request by the Air Force for use on government projects. It is assumed that the SWAT ratings would be collected either after-the-fact, via interviews, or during simulated flight scenarios by interrogating the crewman via voice communications.

o   NASA Task Load Index (TLX) -- this subjective rating technique has been used largely by its developers at NASA Ames. Upon request, NASA Ames supplies PC-based software to implement TLX data collection and analysis for use on government projects. It is assumed that the TLX ratings would be collected in a similar manner to the SWAT.

o   Behavioral measures from flight-related tasks -- e.g., control stick movements (frequency, magnitude, timing in response to distinguishable events), time and nature of manual responses (on appropriate switches, keyboards, pushbuttons) after an event that warrants a response.

o   Event-related potential (ERP) amplitude and latency -- derived from Electroencephalogram (EEG) recordings, ERPs are voltage fluctuations recorded from the scalp that are time-locked to external stimuli. As illustrated by the empirical research accomplished on the present project (see Section 2), transient ERPs are usually characterized by the amplitude, latency from stimulus onset, and scalp distribution of the various component peaks in the waveform. Our work and others (see e.g., Donchin, et al., 1986) have related ERPs to mental workload and other cognitive constructs. It is difficult with conventional electrodes to record EEG data in-flight or in some simulator settings, because the scalp electrodes interfere with the flight helmet or other equipment. However, this should not be a problem in the ACFS simulator, where the aircrew need not wear helmets or life support gear. Efforts are underway at several sites to integrate recording electrodes in flight helmets and to reduce contaminating electrical artifacts that can occur in naturalistic settings; therefore, it may soon be feasible to accomplish high-quality scalp recordings on a routine basis in even the most demanding environments.

o   Blink frequency and duration -- derived from Electrooculogram (EOG) recordings with amplifiers specified herein. Blink rate increases appear to reflect the deterioration in attention and performance which occurs over a prolonged task (Beideman & Stern, 1977; Bauer, et al., 1985). Additionally, Goldstein, et al., (1985) reported that blink durations generally increase with time on task. As workload increases, blink rates decrease and the latency of the blink, after presentation of the stimulus of interest, increases (Bauer, et al., 1987). The pattern of these results are consistent with the notion that as visual information processing demands increase, the eye reflects the subject's attempt to take in more visual input.

o   Heart rate and vagal tone -- derived from Electrocardiogram (EKG) recordings with amplifiers specified herein. Heart rate increases during periods of increased workload, for example during take-offs and landings, have been reported (Roscoe, 1978, 1982; Hart, et al., 1984); but others have not found heart rate to be sensitive to cognitive aspects of simulated flight (e.g., Casali & Wierwille, 1983). More consistent relationships with workload have been reported for heart-rate variability. The typical finding has been that with increased attention and workload, heart-rate variability decreases (e.g., Sayers, 1975; Egelund, 1985; Veldman, et al., 1985). The inter-beat intervals of the heart define a complex time series that reflects a variety of influences on the heart. Respiratory sinus arrhythmia (RSA), the speeding and slowing of the heart beat with breathing, is one such influence that is mediated by the brain. Vagal tone provides a pure estimate of RSA which avoids the statistical assumptions of other procedures, such as spectral analysis, for analyzing heart-rate variability (Porges, 1985).

o   Voice stress -- In recent years, there has been considerable interest in the ability to detect operator workload and stress levels using speech and vocal patterns (e.g., Chambers, et al., 1983; Mosko, et al., 1983; Peckham, 1979; Peckham, 1980). Much of this activity has stemmed from the development of voice recognition systems capable of replacing keyed data entry and manual control of complex platforms in operational settings. In these environments a number of variables challenge the accuracy of voice

recognition capabilities — e.g., acoustic noise, vibration, feedback techniques, training strategies, speech pattern access, response time, vocabulary size and characteristics of particular populations of users.

7.4 Functional Characteristics of the Workload Assessment System

The workload assessment system envisioned here will offer the design engineer or behavioral scientist the following capabilities:

o   Ability to transduce and amplify a user-selectable number of physiological signals from a system operator.

o   Ability to digitize these analog waveforms at a user-selectable sampling rate and store them via direct memory access.

o   Pattern recognition routines to detect artifacts in the physiological recordings (e.g. due to electrical noise from the environment or contaminating electrical potentials from the subject) and state-of-the-art, adaptive algorithms to remove these artifacts from the data.

o   User input facilities for entering subjective ratings of workload.

o   Ability to communicate via digital input/output, both serial and parallel, with a host computer that is running a simulation or operational system in order to acquire behavioral and platform performance data; the system would accommodate a communication protocol similar to the one that was developed in the present project for the VAX 8600 to convey information to the MINC system at NASA Langley; the specified workload assessment system would, of course, be used instead of the MINC.

o   Ability to display an incoming stream of data on a color graphics monitor so that an experimenter can monitor the data acquisition process; the displayed data might be "raw" waveforms, running average waveforms, continuous behavioral indices such as stick outputs, or annunciation of certain events detected by the host system software.

o   Ability to compute derived values of the workload indices mentioned
     earlier; for example, one might wish to compute vagal tone from EKG, to
     compute blink frequency from EOG, to compute reaction time from the onset
     of some event in the cockpit until a switch actuation is accomplished by
     the pilot, to compute "tracking accuracy" from flight parameters such as
     altitude (assuming that the simulator can report the altitude profile of an
     ideal flight path), or to calculate Task Load Index workload values based
     on bipolar rating data entered by the subject.

o   Ability to display, correlate and otherwise relate different variables that
     have been recorded at different sampling rates along a common time base;
     for example, if may be of interest to relate altitude during a particular
     phase of flight (a measure of primary task performance assuming that the
     commanded altitude is known) with heart rate and vagal tone, computed from
     the EKG; altitude may have been encoded at 5 Hz while the EKG was sampled
     at 500 Hz; the workstation would allow the operator to specify heart rate
     and vagal tone values to be calculated at a rate of 1 per every 10 seconds
     and the altitude data to be collapsed (e.g. by averaging) into
     corresponding 10 second epochs, so that a point by point comparison can be
     made graphically or statistically.

o   Ability to display, on a color graphics monitor, raw or derived values of
     groups of variables on an X-Y plot with the X axis being a common time
     base; ability to display appropriate derived variables on bar graphs, X-Y
     scatter plots, or stem and leaf displays.

o   Ability to sort, merge, average or otherwise combine data across flights,
     sessions or subjects.

o   Ability to apply descriptive and inferential statistics to the data; the
     system should include a general purpose statistical package.

o   Ability to apply knowledge-based rules to the data reduction and
     interpretation process; this expert system capability might scan the
     incoming data for particular meaningful patterns or might act as an
     interpretative aid to the data analyst.

o   Ability to perform the above functions in a completely menu-driven manner, without the need for the user to program the workstation; the system will be "user-friendly" with appropriate prompts, diagnostic error messages, and on-line help screens; throughout the software implementation activities a major effort should be made to coordinate the various ongoing development activities and to ensure that a user-friendly man-machine interface results.   Current published human factors design guidelines should be followed.

o   Ability to program the system, using an upper-level command language, to perform functions not initially foreseen.

7.5   System Hardware and Software

A system that meets the functional specifications in the last section was configured based largely on off-the-shelf hardware.   This system specification was detailed in a separate report to NASA which contains information protected by SBIR rights.

7.6   Usefulness of ARD's ANALYZ Package

ARD's proprietary software package for analyzing psychophysiological data (see Section 4) may be useful as part of the present software configuration.   It comprises an extensive set of data analysis and data management routines that can be applied to a wide variety of physiological measures.   Written in FORTRAN, the package has been made largely transportable, user-friendly, and well-documented.   It constitutes a modular, yet well-integrated set of tools for processing ERPs and aspects of other electrophysiological signals.

ANALYZ was initially developed for use on PDP-11 systems running the RT-11 operating system.   ARD has adapted it for use on VAX systems running the VMS operating system.   The present preliminary analyses suggest that the ANALYZ package can be ported, with some modifications, to run in a PC environment (where there are, in fact, some advantages in terms of memory space available for executable code) and that significant parts of the package provide capabilities that would be difficult to implement in existing scientific

software tool-kits. Even if the existing code proves to be difficult to adapt for the present application, the data analytical approaches that it embodies will serve as a useful model for the code to be developed for the present workload assessment system.

7.7 Usefulness of LABPAK Subroutines

The routines that ARD installed on the MINC laboratory computer system at NASA Langley (see Section 6) will also prove useful in the workload assessment system specified here. These routines, written in assembly language (MACRO-11) for that system would be used in applications programs to provide such functions as starting and stopping timers, activating analog-to-digital (A-to-D) conversion, and issuing and receiving digital input and output (I/O). Routines that perform these functions can be implemented in a "canned" manner and then integrated, by the sophisticated programmer, into main programs that address specific experimental goals. The availability of such a package of "low-level" functional subroutines allows the programmer of such main programs to work in a higher level language, without worrying about the need to program the device handlers for each experiment. While this package of subroutines would not be compatible with the hardware architecture of the envisioned workload assessment system, they serve as a useful model for a functionally similar set of routines that should be developed for the specified system.

7.8 Knowledge-based Capabilities

There are two aspects to the expert knowledge that could be brought to bear on the present workload assessment application. On the one hand, there should be pattern recognition algorithms that monitor the incoming data in real-time, to supplement the expert observer, and detect the occurrence of patterns of interest. Depending on their significance, these detections could trigger a message to the user or simply result in the data being tagged for later perusal. In addition, there should be software tools that make it easy for the expert data analyst to retrieve data after-the-fact and to process it in various "intelligent" ways. At this stage of analysis, knowledge-based information could be implemented to direct the automated computation of certain derived measures or to aid the data analyst by suggesting interpretations based

on patterns in the data.

The data should be labelled in a meaningful way such that the expert can rapidly get to the parts of the data base that he is likely to find most informative, and the system should offer strong interactive graphic capabilities for inspecting these segments of the data. Certain ways of quantifying the data could be automated and could be used to suggest possible interpretations to the data analyst. For example, difference waveforms could be calculated between two conditions of interest, or multiple regression and analyses of variance could be performed to examine the effects of different experimental manipulations. Although there is considerable debate as to the feasibility or desirability of carrying this automated inference capability to the point of inferring the functional significance of the data, one can make a strong case for the advisability of attempting to automatically infer trends and interrelationships in the data themselves, as an aid to the human observer.

The sort of artificial intelligence required to implement such capabilities suggests the use of expert system techniques, to compare a set of observables against a set of possibilities, to make probabilistic judgments about the results, and to present the inferences and possible courses of action that follow from a given set of input conditions. However, underlying this higher level decision-making structure, and working in conjunction with it, must be a set of software tools that are specific to the nature of the data being processed. These software tools would provide means of quantifying the electrophysiological phenomena in terms of derived measures, abstracting and trending these measures over time, and recombining them in various ways to reveal their interrelationships. Such capabilities are offered by a number of existing expert system "shell" programs.

7.9 Potential Applications of the Specified System

The workload assessment system that we envision should have wide-ranging usefulness, both for military and non-military applications. While the system we have specified responds directly to the needs of NASA for better workload assessment tools in the ACFS environment and others, the system is by no means specific to any one environment, nor for that matter to cockpit performance

issues. Thus the system recommended here can be readily used with different front-end data acquisition systems in a variety of environments. While the instrumentation that supports this data acquisition may be system-specific, the workload measures that are to be derived are generic, as are the data manipulation and analysis capabilities that the engineer requires at the workstation.

Although the system specified here constitutes a finished system for many workload assessment applications, there are several ways in which it could be expanded with future R & D. Such future development could include the following objectives:

o   To develop interfaces with the next generation(s) of in-flight recording instrumentation; such instrumentation is likely to include data bus recorders with solid-state memory or optical disk-based bus recorders.

o   To develop an in-flight recording package that would provide selective (i.e., "smart") recording of physiological parameters.

o   To develop a microprocessor-based unit for soliciting and encoding pilot's subjective ratings in-flight or during simulated scenarios. In order to be maximally non-intrusive, this unit might make use of voice synthesis and voice recognition algorithms.

o   To develop additional software modules for the ground-based workstation, in order to support "off-line" testing of pilots with batteries of performance tests such as the Criterion Task Set (see e.g., Schlegel, et al., 1986; Shingledecker, 1984); such computerized test batteries may prove useful for predicting individual differences in susceptibility to high workload conditions or for developing benchmark performance measures for comparison with those obtained in simulator or operational settings.

8.0 CONCLUDING SUMMARY

The present Phase II SBIR study was designed to address issues related to scalp-recorded event-related potential indices of mental workload and to examine the feasibility of transitioning ERP indices of mental workload from the laboratory towards operational settings. Such a transition will facilitate the usefulness of psychophysiological measures such as ERPs in the systems engineering process. ERPs obtained from a pilot flying a simulator may complement measures of workload based on behavioral or subjective data and, thereby, support design decisions.

ARD's Phase II project involved five main tasks, which defined a logical progression of ERP measurement technology from a laboratory environment to a simulator-based environment:

1) Two laboratory studies were conducted in order to explore, under controlled conditions, the generality of the encouraging ERP results obtained in the Phase I study. The Phase I findings were generally replicated, even when very different response requirements were imposed on subjects. In addition, two new ERP negativities which were consistently related to workload were discovered. They occurred to a greater extent in the waveforms elicited under high workload conditions than in those elicited under low workload conditions.

2) A "task analysis" of flight scenarios and pilot decision-making in the ACFS was conducted for the purpose of defining events (i.e., displays, messages, alarms) to which the aircrew are exposed during realistic flight scenarios that would be expected to elicit ERPs related to workload. This analysis resulted in a set of candidate cockpit events which can now be validated for use in eliciting ERPs related to workload. In addition, this analysis provided a preliminary test plan for such a validation, as well as some ideas regarding research which could advance the state-of-the-art for recording ERPs under naturalistic conditions.

3) Software was developed to support ERP data analysis; this task included three subtasks -- the upgrade of an existing ARD-proprietary package of ERP data analysis routines, the development of new routines for graphic displays to enhance interactive data analysis, and the development of routines to simulate single-trial ERP data and to systematically compare two alternative single-trial analysis techniques.

4) Working in conjunction with NASA Langley research scientists and simulator engineers, preparations were made for a validation study of ERP measures of workload using the ACFS and laboratory facilities at Langley. A test plan was delineated, incorporating the candidate cockpit events from the task analysis, and provisions were made for implementing ERP measures in the ACFS at Langley. These provisions included the installation of a package of software subroutines on the laboratory MINC computer at NASA Langley and a plan for interfacing the MINC and the ACFS computer.

5) A design specification was developed for a general purpose, computerized, workload assessment system that can function in simulators such as the ACFS or in related operational environments. This system would be of immediate usefulness to NASA in conducting engineering studies in environments such as the ACFS.

9.0 REFERENCES

Aunon, J. I. & McGillem, C. D. (1975). Techniques for processing single evoked potentials. _Transactions of the San Diego Biomedical Symposium_, _14_, 211-218.

Aunon, J. I. & McGillem, C. D. (1979). Detection and processing of individual components in the VEP. _Psychophysiology_, _16_, 71-79.

Bauer, L. O., Goldstein, R., and Stern, J. A. (1987). Effects of information processing demands on physiological response patterns. _Human Factors_, _29_, 213-234.

Bauer, L. O., Strock, B. D., Goldstein, R., Stern, J. A. & Walrath, L. C. (1985). Auditory discrimination and the eyeblink. _Psychophysiology_, _22_, 636-641.

Beideman, L. R. and Stern, J. A. (1977). Aspects of the eyeblink during simulated driving as a function of alcohol. _Human Factors_, _19_, 73-77.

Casali, J. G., & Wierwille, W. W., (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. _Human Factors_, _25_, 623-641.

Chambers, N. C., Brakefield, J. C., Yahiel, D. I., & Fulgham, D. D. (1983). _Stress assessment through voice analysis_ (Report No. AFHRL-TP-83-47). Brooks Air Force Base: Air Force Human Resources Laboratory. (ADA132577/8)

Childers, D. G., Perry, N. W., Fischler, I. S., Boaz, T., Arroyo, A. A. (1987) Event-related potentials: A critical review of methods for single trial detection. _CRC Critical Reviews in Bioengineering_, _14_: 185-200.

Coppola, R., Tabor, R. & Buchsbaum, M. S. (1978). Signal to noise ratio and response variability measurements in single trial evoked potentials. _Electroencephalography and Clinical Neurophysiology_, _44_, 214-222.

Donchin, E. & Herning, R. I. (1975). A simulation study of the efficacy of stepwise discriminant analysis of the detection and comparison of event related potentials. _Electroencephalography and Clinical Neurophysiology_, _38_, 51-68.

Donchin, E., Kramer, A. F. & Wickens, C. (1986). Applications of brain event-related potentials to problems in engineering psychology. In M. G. H. Coles, E. Donchin, & S. Porges (Eds.). _Psychophysiology: Systems, processes, and applications_. New York: Guilford Press.

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive Psychophysiology: The Endogenous Components of the ERP. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.). _Event-related brain potentials in man_. New York: Academic Press, 349-412.

Duncan-Johnson, D. C. & Donchin, E. (1977). On quantifying surprise: The variation in event-related potentials with subjective probability. Psychophysiology, 14, 456-457.

Egelund, N. (1985). Heart-rate and heart-rate variability as indicators of driver workload in traffic situations. In J. F. Orlebeke, G. Mulder, and L. J. P. van Doornen (Eds.). Psychophysiology of cardiovascular control. New York: Plenum Press.

Eggemeier, F. T., Crabtree, M. S., Zingg, J. J., Reid, G. B., & Shingledecker, C. A. (1982). Subjective workload assessment in a memory update task. Proceedings of the 1982 Human Factors Society Annual Meeting, 643-647.

Goldstein, R., Walrath, L. C., Stern, J. A., and Strock, B. D. (1985). Blink activity in a discrimination task as a function of stimulus modality and schedule presentation. Psychophysiology, 22, 629-635

Gopher, D. & Donchin, E. (1987). Workload -- An examination of the concept. In K. Boff & L. Kaufman (Eds.). Handbook of perception and human performance. New York: Wiley & Sons.

Gratton, G., Coles, M. G. H., & Donchin, E. (1989). A procedure for using multi-electrode information in the analysis of components of the event-related potential: Vector filter. Psychophysiology, 26, 222-232.

Harris, E. K. & Woody, C. D. (1969). Use of an adaptive filter to characterize signal-noise relationships. Computers and Biochemical Research, 2, 242-273.

Hart, S. G. (1987). Theory and measurement of human workload. In J. Zeidner (Ed.). Human productivity enhancement: Cognitive processes in system design. New York: Praeger Scientific.

Hart, S. G., Childress, M. E. & Hauser, J. R. (1982). Individual definitions of the term "workload". Proceedings of the 1982 Psychology in the DOD Symposium. USAFA, CO, 478-485.

Hart, S. G., Hauser, J. R., & Lester, P. T. (1984). Inflight evaluation of four measures of pilot workload. Proceedings of the Human Factors Society -- 28th Annual Meeting, 945-949.

Hart, S. G. & Staveland, L. E. (1986). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload, Amsterdam: North Holland Press.

Harter, M. R. & Aine, C. J. (1984). Brain mechanisms of visual selective attention. In R. Parasuraman & D. R. Davies (Eds.). Varieties of attention. Orlando, Fla.: Academic Press, 293-321.

Heffley, E., Foote, B., Mui, T., & Donchin, E. (1985). PEARL II: Portable laboratory computer system for psychophysiological assessment using event related brain potentials. Neurobehavioral Toxicology and Teratology, 7, 399-407.

Horst, R. L. (1988). An overview of current approaches and future challenges

in physiological monitoring. In J. R. Comstock, Jr. (Ed.), Mental-state estimation 1987, NASA Conference Publication 2504, 25-42.

Horst, R. L. & Donchin, E. (1980). Beyond averaging. II: Single trial classification of exogenous event-related potentials using stepwise discriminant analysis. Electroencephalography and Clinical Neurophysiology, 48, 113-126.

Horst, R. L., Johnson, R., Jr. & Donchin, E. (1980). Event-related brain potentials and subjective probability in a learning task. Memory and Cognition, 8, 476-488.

Horst, R. L., Munson, R. C. & Ruchkin, D. S. (1984). Event-related potential indices of workload in a single task paradigm. Proceedings of the 28th Annual Meeting of the Human Factors Society, 727-731.

Horst, R. L., Munson, R. C. & Ruchkin, D. S. (1985). Processing inferences derived from event-related potential measures in a monitoring task. Proceedings of 29th Annual Meeting of the Human Factors Society, 976-980.

Horst, R. L., Ruchkin, D. S. & Munson, R. C. (1986). ERP processing negativities related to workload. In R. Johnson, Jr., R. Parasuraman, & J. W. Rohrbaugh (Eds.). Current trends in event-related potential research (EEG Supplement 40). Amsterdam: Elsevier, 186-190.

Isreal, J. B., Chesney, G. L., Wickens, C. D. & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task paradigms. Psychophysiology, 17, 259-273.

Isreal, J. B., Wickens, C. D., Chesney, G. L. & Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. Human Factors, 22, 211-224.

Johnson, R., Jr. & Donchin, E. (1978). On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology, 44, 424-437.

Kutas, M., McCarthy, G. & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197, 792-795.

Mangun, G. R. R. & Hillyard, S. A. (1987). The spatial allocation of visual attention as indexed by event-related brain potentials. Human Factors, 29, 195-211.

McGillem, C. D. & Aunon, J. I. (1976). Measured characteristics of single evoked brain potentials. Transactions of the San Diego Biomedical Symposium, 15, 225-229.

McGillem, C. D. & Aunon, J. I. (1977). Measurement of signal components in single visually evoked brain responses. IEEE Transactions on Biomedical Engineering, BME-24, 232-241.

Moray, N. (1979). Mental workload: Its theory and application. New York: Plenum Press.

Mosko, J. D., Stevens, K. N., & Griffin, G. R. (1983). <u>Interactive voice technology: Variations in the vocal utterances of speakers performing a stress-inducing task</u> (Report No. NPS55-84-002). Monterey, Calif.: Naval Post-Graduate School. (ADA142554/5)

Naataanen, R. (1982). Processing negativity: An evoked potential reflection of selective attention. <u>Psychological Bulletin</u>, <u>92</u>, 605-640.

Narasimhan, S. V. & Dutt, D. N. (1985). Software simulation of the EEG. <u>Journal of Biomedical Engineering</u>, <u>7</u>, 275-281.

Peckham, J. B. (1979). <u>A device for tracking the fundamental frequency of speech and its application of "strain" in pilots and air traffic controllers</u> (Report No. RAE-TR-79056). Orpington, England: Defense Research Information Center. (ADA085865/4, DRIC-BR-69831)

Peckham, J. B. (1980). <u>The recording and preliminary analysis of a data base for the assessment of "strain" in air traffic controllers, using speech</u> (Report No. RAE-TM-FS-334). Orpington, England: Defense Research Information Center. (ADA093460/4, DRIC-BR-75080)

Pfurtscheller, G. & Cooper, R. (1975). Selective averaging of the intracerebral click evoked responses in man: An improved method of measuring latencies and amplitudes. <u>Electroencephalography and Clinical Neurophysiology</u>, <u>38</u>, 187-190.

Porges, S. W. (1985). Respiratory sinus arrhythmia: An index of vagal tone. In J. F. Orlebeke, G. Mulder, & L. J. P. van Doornen (Eds.). <u>Psychophysiology of cardiovascular control</u>. New York: Plenum Press, 437-450.

Reid, G. B. (1985). The systematic development of a subjective measure of workload. <u>Proceedings of the Ninth Congress of the International Ergonomics Association</u>, 109-111.

Reid, G. B., Shingledecker, C. A., & Eggemeier, F. T. (1981). Application of conjoint measurement to workload scale development. <u>Proceedings of the 1981 Human Factors Society Annual Meeting</u>, 522-526.

Ritter, W., Simson, R. & Vaughan, H. G., Jr. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. <u>Psychophysiology</u>, <u>20</u>, 168-179.

Roscoe, A. H. (1978). Stress and workload in pilots. <u>Aviation, Space, and Environmental Medicine</u>, <u>49</u>, 630-636.

Roscoe, A. H. (1982). Heart-rate as an in-flight measure of pilot workload. <u>Proceedings of the workshop on flight testing to identify pilot workload and pilot dynamics</u> (Report No. AFTEC-TR-82-5). Edwards Air Force Base, CA: Air Force Test and Evaluation Command, 338-349.

Ruchkin, D. S. & Glaser, E. M. (1978). Simple digital filters for examining CNV and P300 on a single trial basis. In D. A. Otto (Ed.). <u>Multidisciplinary perspectives in event-related potential research</u>. Washington, DC: Environmental Protection Agency, EPA-600/9-77-043: 579-581.

Ruchkin, D. S., McCalley, M. G. & Glaser, E. M. (1977). Event-related potentials and time estimation. Psychophysiology, 14, 451-455.

Ruchkin, D. S. & Sutton, S. (1978). Emitted P300 potentials and temporal uncertainty. Electroencephalography and Clinical Neurophysiology, 45, 268-277.

Ruchkin, D. S. & Sutton, S. (1979). Latency characteristics and trial by trial variation of emitted potentials. In J. E. Desmedt (Ed.). Cognitive components in cerebral event-related potentials and selective attention. Karger, Basel, 106-118.

Ruchkin, D. S. & Sutton, S. (1983). Positive slow wave and P300: Association and disassociation. In A. W. K. Gaillard & W. Ritter (Eds.). Tutorials in ERP research - endogenous components. Amsterdam: North Holland, 233-250.

Ruchkin, D. S., Sutton, S., Munson, R. C., Silver, K., & Macar, F. (1981). P300 and feedback provided by absence of the stimulus. Psychophysiology, 18, 271-282.

Ruchkin, D. S., Sutton, S. & Stega, M. (1980). Emitted P300 and slow wave event-related potentials in guessing and detection tasks. Electroencephalography and Clinical Neurophysiology, 49, 1-14.

Sayers, B. McA. (1975). Physiological consequences of informational load and overload. In P. H. Venables & M. J. Christie (Eds.). Research in psychophysiology. New York: Wiley, 95-124.

Schlegel, R. E., Gilliland, K., & Schlegel, B. (1986). Development of the Criterion Task Set performance data base. Proceedings of the 30th Annual Meeting of the Human Factors Society, 58-62.

Sexton, G. A. (1983). Crew Systems and Flight Station Concepts for a 1995 Transport Aircraft (Report No. 166068). Hampton, VA.: NASA Langley Research Center.

Sexton, G. A. & Needles, C. K. (1982). Mission analysis documents for development of a 1995 flight station (Report No. LG82ER0054). Marietta, GA: Lockheed-Georgia Company.

Shingledecker, C. A. (1984). A task battery for applied human performance assessment research (Report No. AFAMRL-TR-84-071). Wright Patterson Air Force Base: Armstrong Aeromedical Research Laboratory.

Squires, K. C. & Donchin, E. (1976). Beyond averaging: The use of discriminant functions to recognize event-related potentials elicited by single auditory stimuli. Electroencephalography and Clinical Neurophysiology, 11, 449-459.

Squires, K. C., Wickens, C., Squires, N. K. & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the critical event-related potential. Science, 193, 1142-1146.

Veldman, J. B. P., Mulder, L. J. M., Mulder, G., van der Heide, D. (1985). Attention, effort and sinus arrhythmia: How far are we? In J. F. Orlebeke, G. Mulder, & L. J. P. van Doornen (Eds.). Psychophysiology of cardiovascular

control. New York: Plenum Press, 407-424.

Wastell, D. G. (1977). Statistical detection of individual evoked responses: An evaluation of Woody's adaptive filter. _Electroencephalography and Clinical Neurophysiology_, _42_, 835-839.

Wastell, D. G. & Kleinman, D. (1980). N1-P2 correlates of reaction time at the single-trial level. _Electroencephalography and Clinical Neurophysiology_, _48_, 191-196.

Weinberg, H. & Cooper, R. (1972). The recognition index: A pattern recognition technique for noisy signals. _Electroencephalography and Clinical Neurophysiology_, _33_, 608-613.

Wennberg, A. & Zetterberg, L. H. (1971). Application of computer-based model for EEG analysis. _Electroencephalography and Clinical Neurophysiology_, _31_, 457-468.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.). _Attention and performance VIII_. Hillsdale, NJ: Erlbaum, 239-258.

Wickens, C. D., Isreal, J. & Donchin, E. (1977). The event-related cortical potential as an index of task workload. _Proceedings of the 21st Annual Meeting of the Human Factors Society_.

Wickens, C. D., Kramer, A., Vanasse, L. & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. _Science_, _221_, 1080-1082.

Woody, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. _Medical & Biological Engineering_, _5_, 539-553.

Zetterberg, L. H. (1969). Estimation of parameters for a linear difference equation with application to EEG analyses. _Mathematical Biosciences_, _5_, 227-275.

Zetterberg, L. H. & Ahlin, K. (1975). Analog simulator of EEG signals based on spectral components. _Medical & Biological Engineering_, 272-278.

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-4240		

4. Title and Subtitle		5. Report Date
Brain-Wave Measures of Workload in Advanced Cockpits: The Transition of Technology From Laboratory to Cockpit Simulator		June 1989
		6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Richard L. Horst, David L. Mahaffey, and Robert C. Munson	
	10. Work Unit No.
	324-02-00

9. Performing Organization Name and Address	11. Contract or Grant No.
Advanced Resource Development Corporation 9151 Rumsey Road Columbia, Md. 21045	NAS1-18019
	13. Type of Report and Period Covered

12. Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Langley Research Center Hampton, Va. 23665-5225	14. Sponsoring Agency Code

15. Supplementary Notes

Langley Technical Monitor: Alan T. Pope
SBIR Phase II Final Report

16. Abstract

The present Phase II SBIR study was designed to address issues related to scalp-recorded event-related potential (ERP) indices of mental workload and to transition this technology from the laboratory to cockpit simulator environments for use as a systems engineering tool. The project involved five main tasks: 1) Two laboratory studies confirmed the generality of the ERP indices of workload obtained in the Phase I study and revealed two additional ERP components related to workload. 2) A "task analysis" of flight scenarios and pilot tasks in the Advanced Concepts Flight Simulator (ACFS) defined cockpit events (i.e., displays, messages, alarms) that would be expected to elicit ERPs related to workload. 3) Software was developed to support ERP data analysis. An existing ARD-proprietary package of ERP data analysis routines was upgraded, new graphics routines were developed to enhance interactive data analysis, and routines were developed to compare alternative single-trial analysis techniques using simulated ERP data. 4) Working in conjunction with NASA Langley research scientists and simulator engineers, preparations were made for an ACFS validation study of ERP measures of workload. 5) A design specification was developed for a general purpose, computerized, workload assessment system that can function in simulators such as the ACFS.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Event-related potential Mental workload cockpit simulator Human performance	Unclassified — Unlimited Subject Category 54

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	196	A09